

# THIS WEEK

## EDITORIALS

**COLLABORATION** International links risk creating two-tier research at home **p.536**

**WORLD VIEW** Isaac Newton, scientific uncertainty and football fans **p.537**



**SUBDUCTION** Plate tectonics can give birth to precious stones **p.538**

## Overtaken by events

*Despite the small number of entries, the genomics X prize is to be commended for attempting to push the boundaries of DNA sequencing technology.*

The sparse list of contenders for the Archon Genomics X Prize shows how far sequencing technology has come — and how far it still has to go. Barring any late surprises, only two teams will have signed up by the registration deadline of 31 May to compete for the US\$10-million prize for the first to sequence 100 centenarians' genomes in 30 days or less at a cost of \$1,000 per genome (see page 546). The sequences must have no more than one in a million errors, be 98% complete and have correct haplotype phasing — a determination of which parent contributed each portion of a chromosome. It is not possible at present to meet this combination of goals with any single technology, but that does not explain why so few are reaching for it.

One reason why more teams are not lining up for the prize is that the promise of a genomic medical revolution is not being stalled by any lack of data. At genetic-medicine conferences, such as the University of California's OME 2013 precision-medicine conference, held on 2 and 3 May in San Francisco, or the Big Data in BioMedicine conference held from 22 to 24 May at Stanford University in California, you will hear the same refrain. "We have more data right now than we know what to do with." Figuring out how to interpret genetic data — and, more crucially, how to prove their value to patients and health-care systems — is the most pressing challenge in genomics today. Researchers can already sequence the protein-coding regions of a genome for less than \$1,000. Getting more data on regions of the genome that they do not yet know how to interpret will not help to advance the goal of proving the medical worth of big data.

Interpretation and analysis — making sense of the data — is now the real prize. Hence the launch of a spate of bioinformatics challenges (see page 547) as researchers compete to surmount that hurdle. They include Sequence Squeeze, a contest to develop the best sequence-data-compression algorithm; the Assemblathon, for the best program to assemble a genome sequence from scratch; the DREAM Challenges to analyse and predict biological interactions among gene products; the CLARITY genome-interpretation challenge; and contests at the annual Beyond the Genome meeting. Michael Schatz at the Cold Spring Harbor Laboratory in New York, who has curated many of these contests, is planning more challenges this year, including one at Cold Spring Harbor later this autumn. Bioinformatics contests have the advantage that they do not require physical manufacturing infrastructure, so they are more accessible to more would-be solvers around the world.

There are other reasons why the genome X prize is a harder sell than other X prizes. The sequencing field is much more mature than were other industries that have been the focus of successful X prizes. Whereas there was no space-tourism industry before dozens of teams competed for the Ansari X Prize in 2004, for instance, there is already a thriving commercial market for sequencing. So any company that could meet the goals laid out in the prize already has its incentive — and it would be worth a lot more than \$10 million. The value of the

market leader in sequencing, Illumina of San Diego, California, is currently \$8.8 billion.

And it is very unlikely that anyone other than a well-financed lab or large company could attempt the current challenge. That also sets it apart from other competitions — the Google Lunar X Prize, for example — in which teams of professionals or even hobbyists can make a respectable showing. The thriving do-it-yourself biology movement, by contrast, cannot mount a credible challenge to the large life-sciences

**"The promise of a genomic medical revolution is not being stalled by any lack of data."**

companies. The attempt is even beyond most biotech start-ups. The UK-based biotechnology company Oxford Nanopore, for example, which is trying to commercialize a promising technology pioneered by highly respected researchers, has raised at least \$150 million in grants and investment since 2008 — but has yet to show that its technology can be used to sequence a complete human genome.

That is not to say that the genomics X prize does not matter. The X Prize Foundation should be commended for revising the challenge, initially laid out in 2006, as the field evolved. It has also done a valuable service by working for two years with many partners, including *Nature Genetics*, to outline a judging scheme that can independently assess the quality and accuracy of a genome sequence and that is agnostic about the sequencing technology used. The foundation deserves kudos for prompting the field to reach farther; if past history is any guide, genomicists will reach that goal sooner than now seems possible. ■

## Still less equal

*Japan's government must stick by its promise to help women's careers to prosper.*

In 2010, there were 1,552 children waiting to get places in child-care centres in Yokohama, by far the highest number of any city in Japan. Over the next three years, the city's (female) mayor, Fumiko Hayashi, spent 37 billion yen (US\$362 million) on building new infrastructure, including 144 childcare centres. Now the waiting list is zero.

Many female scientists, as well as women working in other sectors, celebrated the news. They know that help with child-rearing responsibilities is essential for a mother to have a successful career. But even better were the reverberations, which reached all the way up to the prime minister, with an indication that change might become more widespread.

On 20 May, after touring one of the childcare centres with Hayashi, Prime Minister Shinzo Abe said that the "Yokohama model" should

be applied across the country. In fact, an economic growth strategy report released in April called for childcare capacity to be increased by 400,000 nationwide.

Why the sudden focus on such a progressive issue from a man who refuses even to consider a popular amendment to Japanese law that would allow the imperial line to pass through female members of the family? The momentum for change seems to be coming largely from a recognition of economic and demographic realities. Japan's population, and its labour force, are shrinking, and its economic competitiveness is faltering. The country has been frittering away a resource that could help to meet these challenges. And the call is getting louder.

In 2010, for example, the investment bank Goldman Sachs, headquartered in New York, released a widely cited report on 'Womenomics' in Japan stating that closing the country's gender gap — bringing the employment rate of women up to that of men — would increase the workforce by 8.2 million and boost gross domestic product by 15%.

Such figures have grabbed the attention of business leaders. Companies are more likely to allow flexible working hours for mothers, for example; government officials are making women's issues part of their political platforms; and 'Abenomics', as the prime minister's aggressive financial and economic plans are called, seems to be endorsing Womenomics (see page 548). But will the right improvements be made?

The problem facing women in Japan has nowhere been as glaring as in science and engineering. The country has taken steps to remedy the situation, but there is much more to be done. In 2006, for example, the science and education ministry brought in a programme to improve support systems for female scientists. Renewed in 2011, the programme has so far supported projects at 88 universities and research institutions.

These projects led to increased on-campus childcare facilities and practical support for scientists with children. Most universities now have at least one childcare centre. Another programme, launched by the science ministry in 2009, and called Supporting Positive Activities for Female Researchers, has funded 5-year programmes at 12 universities

to increase the number and promotion rate of female faculty members. These 'affirmative action' programmes have ramped up the number of female scientists, especially at higher-level positions.

And such policies have created momentum. At institutions that implemented the plans, the proportion of women among the research staff grew from 12.5% in 2005 to 15.4% in 2011. Nationwide figures have risen from 11.9% to 13.8% over the same period. But this

increase is too slow. Japan still has the lowest number of females as a percentage of total science researchers in developed countries, according to figures collected by the science ministry. A new innovation policy from the Council for Science and Technology Policy, currently in draft form, calls for the percentage of female new hirings at universities and research institutes to reach 30% by 2016. It is an ambitious goal, considering that this figure now is just over 20%.

Ensuring equal opportunities for women scientists is a global problem. But Hisako Ohtsubo, a molecular biologist at Nihon University near Funabashi, says that Japan is more than 25 years behind some countries, noting that the US National Science Foundation brought in such policies in the 1980s. To catch up, the government should expand its programmes, which have so far been short-term and on too limited a scale, and ensure that they encourage women not only to engage in science but also to fight for leadership positions in the scientific community.

With competition growing, especially from China and South Korea, Japan's position as a scientific power is no more secure than its status as an economic power. The Japanese government seems to be waking up to women's potential importance to the economy. Making the most of their talent could be just as potentially transforming for Japanese science. The Abe government needs to stick by his promises and take the targets seriously. ■

**"The Japanese government seems to be waking up to women's potential importance to the economy."**

## Without borders

*Increasing scientific globalization is welcome, but could compromise national efforts.*

"It might be a good idea," the US journalist Bill Vaughan once suggested, "if the various countries of the world would occasionally swap history books, just to see what other people are doing with the same set of facts."

History is famously written by the winners, and does tend to huddle inside national boundaries. Science is different, or so researchers like to believe. It is an international process. Facts are shared. Differences are ironed out. Scientists know what other people are doing. Correct?

Yes and no. As an impressive analysis of some 25 million research papers in a Comment on page 557 shows, international collaboration plays an increasing part in twenty-first-century research. The growth in scientific output from the United States and western Europe in the past three decades, for example, is entirely down to international collaboration. The number of journal articles that feature authors entirely from a single country in those regions has remained the same.

Only because of global links can the UK government claim, as it did in a 2011 report, that the country creates "14 per cent of the world's highly cited output" with just "4 per cent of the world's gross expenditure" on research and development. Most of those high-impact papers include the contribution of a foreign scientist. In fact, in 2010, the number of 'British' papers produced with help from abroad exceeded those made entirely in Britain for the first time.

According to the US National Science Foundation, almost

one-quarter of global research articles in 2010 featured authors from more than one country, up from 10% in 1990. The average number of authors has doubled since 1980 and now stands at 4.5.

Although *Nature* welcomes the global reach of science, we have previously pointed out that internationalization will rub up against some natural boundaries. National pride and prestige matter, not least because science is still mostly funded and managed on a national basis.

Last October, in a special issue on the globalization of science, we said: "Mobility cannot stretch infinitely: relationships, families and quality of life put limits on how much researchers want to travel, and for how long" (*Nature* 490, 309–310; 2012). And we pointed out that collaboration could blur the borders of national priorities, especially for countries that are just beginning to develop their science bases.

The science bases of some of the bigger developing economies are coming along just fine, thank you. As the Comment points out, most of the scientific growth in China, Brazil and South Korea is driven by domestic work. And quality is rising. The citation impact of more than 10% of China's domestic research is greater than twice the world average.

What does this mean? A change in international dynamics for starters, as Jonathan Adams points out in the Comment: "The older economies can no longer rely on the best foreign researchers to come and visit."

There are opportunities galore in this new world, but Adams also highlights some of the threats. The best institutions in different countries already tend to work together the most. Others can find it hard to join this global competition. "There is a growing divide between international and domestic research," Adams notes. "This will influence each nation's ability to draw on the global knowledge base, and could in turn compromise national scientific wealth." It is not just between nations that science must be shared, but within their borders too. ■

➔ **NATURE.COM**  
To comment online,  
click on Editorials at:  
[go.nature.com/xhunjv](http://go.nature.com/xhunjv)



## Tell fans definitive calls are an impossible goal

The electronic eyes that will watch football goal lines this summer may settle arguments, but they still depend on probability, says Nic Fleming.

Almost three years ago, the English footballer Frank Lampard scored a goal in a World Cup quarter final against Germany. The ball hit the crossbar and bounced down way over the line, before it was cleared by a grateful goalkeeper. The referee and his assistants failed to award the goal, because they could not be sure the ball had crossed the line.

Every fan of football (apologies to US readers but I refuse to use the term soccer) has their own story of a goal that should or shouldn't have been given. Those in charge of the game have finally bowed to public pressure and next month's Confederations Cup in Brazil will see goal-line technology used in an international tournament for the first time. In August, a similar system will be introduced into the English Premier League, the richest and highest-profile league in the world. The companies involved are already discussing how their machines could be applied to help football referees make offside, handball and other decisions. Sports officials in the United States and Canada are watching with interest.

Yet the introduction of goal-line technology to football is likely to perpetrate a mass deception on television viewers. It will miss a huge opportunity to educate people about the role of uncertainty in science. And it will exacerbate the approaching danger of fake computer-generated video footage.

Have you seen how a tennis player's challenge of a close line call is adjudicated by the Hawk-Eye system? I confess that until recently I was one of the vast majority who believed they were watching an accurate replay of reality. In fact, the system combines multiple high-speed camera images to estimate the ball's trajectory and position when it hits the ground. Nothing in the graphics indicates to viewers the inevitable uncertainty and potential for error in that estimation. As it stands, the same will apply to goal-line replays.

Fans are likely to share a misunderstanding that apparently reaches right to the top of the game. Announcing the introduction of Hawk-Eye to football in April, the Premier League's chief executive, Richard Scudamore, said that the technology would reveal "definitively" whether the ball had crossed the line.

Scudamore may not have read the 1726 edition of Isaac Newton's *Philosophiæ Naturalis Principia Mathematica*, which stresses the role of uncertainty in the scientific method. Measurement involves error. Results are accompanied by confidence limits and levels. With public opinion key to decisions on topics such as climate change, nuclear power and genetic modification, calls to spread awareness of uncertainty and probability in science are increasing, even if they are often drowned out by the comforting simplifications of certainty that characterize mainstream reporting.

In 2011, sociologists at Cardiff University, UK, published a paper on the use of Hawk-Eye in sports (H. Collins & R. Evans *Public Understanding of Science* <http://doi.org/frmw94>; 2011). The system introduced a "false transparency" in tennis, they said, and should be supplemented with visual error bars or numerical statements of confidence in the screened Hawk-Eye graphics.

Cricket fans already understand this. In leg-before-wicket decisions (sorry again US readers — look it up) there is a 'zone of uncertainty' where Hawk-Eye admits that it cannot be certain whether or not the ball would have hit the stumps. It defers to the judgement of the human umpire and this is acknowledged to television viewers with an 'on-field call' message. Football fans will be denied such information.

Both Hawk-Eye and the (separate) GoalControl technology that will be used in the Confederations Cup train seven high-speed cameras

on each goal. Computer software then combines the two-dimensional images into three-dimensional representations of the ball and its position. Both systems can probably provide more accurate decisions than human referees. Hawk-Eye claims a minimum accuracy of  $\pm 6$  millimetres whereas GoalControl claims  $\pm 5$  millimetres. These claims cannot be verified as the International Federation of Football Associations refuses to release the test results, beyond confirming that both systems meet its minimum requirement of  $\pm 3$  centimetres in tests of balls moving at 70 kilometres per hour.

The public is largely aware that still images can be airbrushed, improved and otherwise manipulated. Many media organizations have strict

codes governing digital tinkering. Earlier this year, North Korea was seemingly caught out cutting and pasting extra hovercrafts into an image to boost its apparent military muscle.

Moving images can seem more realistic. Yet we are rapidly approaching a time when computer-generated moving images will look as if they have been filmed. We will be able to create entirely faked video footage. This is useful to create an impressive mountain backdrop for a film. But what about when political propagandists or those with commercial motives start to exploit such technology?

If video footage is to retain credibility, we need greater transparency about visual representations of events, so we can distinguish replays from reconstruction. On-screen honesty on the limitations of goal-line technology would do this, and put a great many people in the picture about science and uncertainty. Football is only a game but it is a good place to start. ■

Nic Fleming is a freelance science and technology journalist in London and a Sheffield United fan.  
e-mail: [nicfleming106@hotmail.com](mailto:nicfleming106@hotmail.com)

THE INTRODUCTION  
OF GOAL-LINE  
TECHNOLOGY  
TO FOOTBALL  
IS LIKELY TO  
PERPETRATE A MASS  
DECEPTION.

➔ NATURE.COM  
Discuss this article  
online at:  
[go.nature.com/3upjdk](http://go.nature.com/3upjdk)



# RESEARCH HIGHLIGHTS

Selections from the  
scientific literature

## NEUROSCIENCE

### Bee exploration mechanism

Honeybees orient using similar molecular pathways to many vertebrates.

Gene Robinson and Claudia Lutz at the University of Illinois at Urbana–Champaign identified a protein in honeybees (*Apis mellifera*) similar to the transcription factor Egr1, which is expressed in the brains of vertebrates such as rodents during the exploration of new environments.

Egr was upregulated only in regions of the bee brain called mushroom bodies — which integrate sensory input and process memories — and only as the bees learned to orient in unfamiliar surroundings. This was true of both young bees that had previously never left the hive and experienced foragers placed in a new environment. The results demonstrate the deep evolutionary conservation of Egr-related molecular pathways in experience-dependent learning, say the authors.

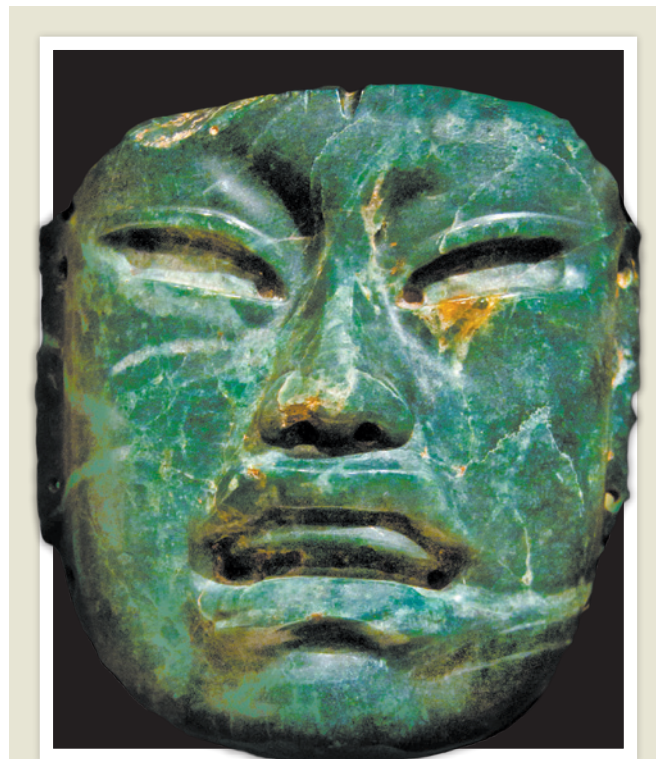
*J. Exp. Biol.* 216, 2031–2038 (2013)

## GENOMICS

### Irish-famine pathogen decoded

Researchers have sequenced the genome of the microorganism that devastated the Irish potato crop in the 1840s — the first time the genome of a historical plant pathogen has been decoded.

Kentaro Yoshida at the Sainsbury Laboratory in Norwich, UK, and his colleagues identified the strain of *Phytophthora infestans*, which causes potato late blight, from nineteenth-century



## GEOLOGY

### Gemstones from the deep

The precious stones jade and ruby can be used to identify the sites of ancient collisions of tectonic plates.

Robert Stern at the University of Texas at Dallas and his colleagues suggest that these two substances could be called “plate tectonic gemstones”. Jadeite, a type of jade, forms where one tectonic plate plunges beneath another. Fluids that rise from the diving slab of oceanic crust condense to form the gemstone (pictured in an Olmec mask from southern Mexico — the site of one such subduction zone). By contrast, ruby forms where continental crust rich in aluminium collides, as in Southeast Asia.

These stones should be recognized not only for their beauty but also for what they reveal about their tectonic setting, the authors say.

*Geology* <http://dx.doi.org/10.1130/G34204.1> (2013)

dried leaves. The researchers sequenced DNA from 11 historical specimens, which had been stored in herbaria, and from 15 modern strains of *P. infestans*. The famine strain was closely related to another strain that is still prevalent around the world,

and the researchers suggest that the two strains diverged in the early 1800s. The famine strain may now be extinct, the authors say.

*eLIFE* <http://dx.doi.org/10.7554/elife.00731> (2013)

For a longer story on this research see [go.nature.com/qg2aov](http://go.nature.com/qg2aov)

## HUMAN EVOLUTION

### Footprints reveal hominin size

Fossil footprints indicate that hominins were already as large as modern humans by 1.52 million years ago.

Undamaged fossil skeletons from that time are rare, and so determining characteristics such as the size and walking speed of human ancestors has been challenging. Brian Richmond and Heather Dingwall at George Washington University in Washington DC and their colleagues measured foot size and stride length from fossil footprints of seven individuals — which were probably *Homo erectus* or *Paranthropus boisei* — discovered in northern Kenya. To translate these measurements into physical attributes such as stature, body mass and walking speed, the researchers studied the relationship between body dimension and gait in habitually barefoot modern adults from Kenya. The authors were then able to infer that the size of these hominins was comparable to that of modern humans.

*J. Hum. Evol.* 64, 556–568 (2013)

## CHEMISTRY

### Metabolites, cell by cell

Single-cell measurements are revealing how individual yeast cells react to environmental and genetic challenges.

Advances in mass spectrometry, which identifies individual compounds in complex mixtures, have given researchers the chance to compose cell-by-cell portraits of metabolism. Renato

PETER HORREE/ALAMY



# SEVEN DAYS

The news in brief

## EVENTS

### MRI spy charges

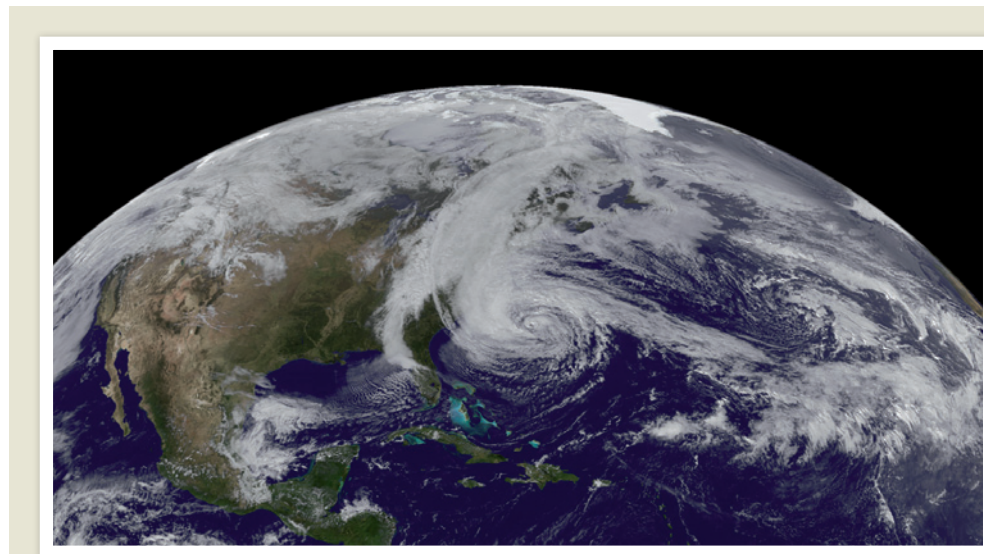
Three Chinese researchers working in the United States have been charged with commercial bribery, announced the Federal Bureau of Investigation and the US Attorney's Office in Manhattan last week. Officials say that radiology professor Yudong Zhu recruited Xing Yang and Ye Li to help him develop magnetic resonance imaging (MRI) technology at New York University's Langone Medical Center. While funded by a multimillion-dollar grant from the US National Institutes of Health, the three allegedly took payments from a rival medical-imaging company and a research institute in China in exchange for confidential information about their research.

### Spain unrest

In response to steep budget cuts, Spanish scientists, university officials and science advocates have called for protests in defence of science in the country. In a 21 May document, the Open Letter for Science movement urged participation in demonstrations on 14 June against what it described as the ruin of the Spanish research system. Spain has reduced its state budget for science by roughly 40% since 2009, and research funds for 2013 are being delayed. The group asks the government to boost science funding and implement measures to retain the country's scientific talent.

### Lobbying lull

Citing huge advances under US President Barack Obama in the availability of human embryonic stem cells for US-funded researchers, one key advocacy group says that its work is done. The Coalition



ROBERT SIMMON/GOES PROJECT SCIENCE TEAM/NASA/NOAA

## Eye of the storm

The National Oceanic and Atmospheric Administration (NOAA) has predicted an extremely active Atlantic hurricane season, six months after Hurricane Sandy devastated the US northeast (pictured). In its forecast released on 23 May, the agency says there is a 70% chance that 7–11 hurricanes, including 3–6 major ones, will form between 1 June and 30 November.

“There are no mitigating factors that we can see that would suppress the activity,” says Gerry Bell, lead hurricane forecaster for NOAA. But the agency says that a new supercomputer and an updated forecasting model — fed with real-time Doppler radar data collected by aeroplanes that fly through storms — should improve the accuracy of its hurricane tracking.

for the Advancement of Medical Research, based in Washington DC, announced its dissolution on 21 May, after lobbying for federal funding for the controversial research since 2001. The news comes nearly five months after the US Supreme Court refused to hear a legal challenge that attempted to stop government support of the research (see [go.nature.com/5jegit](http://go.nature.com/5jegit)).

### Deep shocks

Eastern Russia was struck by a deep, magnitude-8.3 earthquake off the Kamchatka peninsula on 24 May. The epicentre was located 610 kilometres below Earth's surface in the Sea of Okhotsk and about 400 km northwest of the city of Petropavlovsk-Kamchatsky, according to

the US Geological Survey. Data suggest that the quake ruptured a deep fault in the Pacific plate, which is being forced under neighbouring plates. See [go.nature.com/kipwvjv](http://go.nature.com/kipwvjv) for more.

### Remodelled sub

With upgrades complete, the US flagship submersible *Alvin* is en route to Astoria, Oregon, from where it will return to exploring the dark depths of the oceans in December. *Alvin* left the Woods Hole Oceanographic Institution in Massachusetts on 25 May aboard the research ship *Atlantis*, after being taken out of service in 2010 for a US\$41-million makeover. Among its improvements is a new titanium sphere — in which researchers sit during

dives — equipped with extra windows and high-definition cameras. See [go.nature.com/mdgxhy](http://go.nature.com/mdgxhy) for more.

## BUSINESS

### Antibiotics deal

In the first partnership of its kind, the US government will pay GlaxoSmithKline up to US\$200 million to develop drugs to combat antibiotic resistance and bioterrorism, the London-based pharmaceutical giant announced on 22 May. The Biomedical Advanced Research and Development Authority, a federal office established to address public-health emergencies and biodefence threats, will pay the company \$40 million over 18 months and up to

\$160 million more if the contract is renewed over 5 years.

## Sleep aid

On 22 May, advisers to US drug regulators recommended approval of a first-in-class therapy for insomnia, but proposed its use only at low doses. The drug, called suvorexant, is made by Merck, a pharmaceutical company in Whitehouse Station, New Jersey. If approved, it will be the first insomnia drug to block orexin, a neurotransmitter that regulates sleep–wake cycles. Reviewers at the US Food and Drug Administration expressed concern about the side effects seen at higher doses, including suicidal thoughts and narcolepsy-like symptoms.

## PEOPLE

## Shaw prizes

Six scientists in Britain and the United States won this year's Shaw prizes, each worth US\$1 million. Jeffrey Hall at the University of Maine in Orono, Michael Rosbash at Brandeis University in Waltham, Massachusetts, and Michael Young at the Rockefeller University in New York share the prize in life sciences and medicine for work on circadian rhythms. Steven Balbus (pictured) at the University of Oxford, UK, and John Hawley



at the University of Virginia in Charlottesville won in astronomy for work on disks of gas and dust around stars. The mathematics prize went to David Donoho of Stanford University in California, for advances in statistics.

## RESEARCH

## Stem-cell censure

Criticisms of apparent data duplication have tarnished a recent study in *Cell* that reported the creation of human stem-cell lines through cloning. An anonymous online commenter noted apparent duplicated images and scatterplots in the paper by Shoukhrat Mitalipov of the Oregon Health and Science University in Portland, and his colleagues. Mitalipov says that the duplications were innocent mistakes, and is working with *Cell* to correct them. See page 543 for more.

## FACILITIES

## Swiss biotech site

Plans for a major Swiss biotechnology initiative have moved forward with the purchase of a host site, announced on 22 May. The Campus Biotech consortium, backed by billionaires Hansjörg Wyss and Ernesto Bertarelli, has bought the 45,000-square-metre former Merck Serono facility in Geneva, Switzerland, for an undisclosed amount. The group plans to open an institute for bio- and neuroengineering in collaboration with the Swiss Federal Institute of Technology in Lausanne and the University of Geneva. See [go.nature.com/bxaqxp](http://go.nature.com/bxaqxp) for more.

## Making waves

Plans for the world's largest wave-energy farm, off the northwest coast of Lewis in Scotland, were given the all-clear by the Scottish government on 22 May. Edinburgh-based firm Aquamarine Power will install 40–50 wave-energy devices, with an overall power capacity of about 40 megawatts once electricity cables are installed, which will probably be 2017 at the earliest.

## Arctic break-up

Fracturing Arctic ice has prompted Russia to order the evacuation of its North Pole 40 research station, which collects

## COMING UP

### 1–5 JUNE

The interaction between human activities and the environment takes centre stage at the third World Summit on Evolution, held in the Galapagos Islands.

[go.nature.com/reklva](http://go.nature.com/reklva)

### 2–6 JUNE

Updates on the search for habitable worlds and results from the Curiosity Mars rover are among the programme highlights of the 222nd American Astronomical Society meeting in Indianapolis, Indiana.

[go.nature.com/2mqa71](http://go.nature.com/2mqa71)

meteorological observations and monitors pollution. The ice floe that hosts the 16-crew station has broken into six pieces, according to news reports last week. In a statement, the environment ministry expressed concern for staff safety, and noted the possible risk of environmental contamination near Canada's economic zone in the event of a disaster.

## POLICY

## Water priorities

The US Agency for International Development released its first global water and development strategy on 21 May. The foreign-assistance agency already spends about US\$452 million each year on water-related programmes, but said that the new plan “elevates the importance” of water as a priority for the group. From this October, the organization will focus all new water programmes on two goals: improving human health and enhancing food security by providing clean, sustainable water.

➔ [NATURE.COM](http://NATURE.COM)

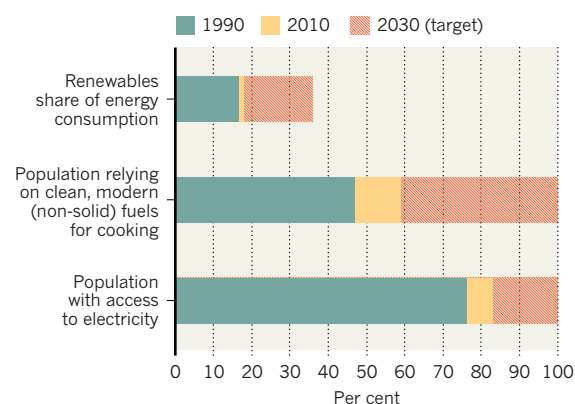
For daily news updates see:  
[www.nature.com/news](http://www.nature.com/news)

## TREND WATCH

The United Nations is leading a push to give the entire world access to electricity and less-polluting cooking fuel by 2030. The agency also aims for renewables to constitute a greater share of energy consumption. But initial statistics released on 28 May in a Global Tracking Framework report (see chart) show that those targets will be hard to reach, mainly because rising populations and economic growth are countering improvements. See [go.nature.com/gyowo6](http://go.nature.com/gyowo6) for more details.

## WIDER ACCESS TO CLEAN ENERGY NEEDED

World falling short of 2030 target to expand global access to electricity and clean household fuels.



# NEWS IN FOCUS

**ENTOMOLOGY** Scientists abuzz about appearance of 17-year cicadas **p.545**

**GENETICS** Tepid turnout for DNA-sequencing prize **p.546**

**JAPAN** Ambitious plan to boost science innovation **p.548**



**NEUROSCIENCE** Karl Deisseroth opens a window on the brain **p.550**

OREGON HEALTH & SCIENCE UNIV.



Shoukhrat Mitalipov, lead author of human stem-cell cloning paper, admits errors.

## STEM CELLS

# Fallout from hailed cloning paper

*Duplicated figures raise debate over expedited publication.*

BY DAVID CYRANOSKI

**H**ow fast is too fast for review of a scientific article? And who has the responsibility to ensure accuracy? Errors found in a widely acclaimed cloning study have rekindled those questions — and sent the lead author and the journal that published it scrambling to assure the world that the problems did not compromise the findings.

The paper, which was published online by the journal *Cell* on 15 May (<http://doi.org/mkn>), reported the creation of human embryonic stem-cell lines from cloned human skin cells. The lines are expected to answer fundamental questions about the way in which cells are reprogrammed and also to have potential

therapeutic applications.

But last Wednesday, after an anonymous online commenter noted three pairs of duplicated images with conflicting labels in the paper, excitement turned to confusion — and a bit of déjà vu. The last time the same feat was claimed — by then Seoul University professor Woo Suk Hwang — duplicate images were noted anonymously and the breakthrough was later debunked. Nobody is claiming more than sloppiness in the present case, and the authors quickly stepped up to put the record straight.

The leader of the study, Shoukhrat Mitalipov, a reproductive-biology specialist at the

Oregon Health and Science University in Portland, acknowledges that the team did use three sets of duplicated images. In two cases that was intentional, but a pair of labels was mistakenly reversed, he says (see <http://doi.org/mnk>).

In the third case, an incorrect scatterplot was included in a table examining the patterns of gene activity in the cloned stem-cell lines — one that had already been used in the paper.

Mitalipov says that corrections will be made. He waves off another criticism raised by the anonymous commenter concerning the surprisingly tight shape of two scatterplots: they just turned out that way, he says.

“The results are real, the cell lines are real, everything is real,” he says.

The editor-in-chief of *Cell*, Emilie Marcus, released a statement in response to the criticisms, defending the paper: “Based on our own initial in-house assessment of issues raised and in initial discussions with the authors, it seems that there were some minor errors made by the authors when preparing the figures for initial submission. While we are continuing discussions with the authors, we do not believe these errors impact the scientific findings of the paper in any way.”

Most researchers seem ready to give Mitalipov the benefit of the doubt, on the basis of his answers and his track record. “The explanations [by Mitalipov] are plausible, but we will have to wait for the results of a thorough investigation,” says Martin Pera, a stem-cell expert at the University of Melbourne, Australia.

Robin Lovell-Badge, a developmental biologist at the Medical Research Council’s National Institute for Medical Research in London, also accepts that the errors resulted just from sloppiness. “I really don’t think in this case it’s any attempt to manipulate. It’s not the data you would want to manipulate, anyway,” he says.

But many also noted that the paper had been published with blazing speed — *Cell* accepted it just three days after receiving it and published it online 12 days later — and questioned whether such rapid publication is good for science. “Whatever the explanation is, it’s amazing that there is another issue with a paper in SCNT [somatic-cell nuclear transfer]. The four-day review process was obviously inadequate,” says Arnold Kriegstein, director of the stem-cell programme at the University of California, San Francisco. ▶

**NATURE.COM**  
For more on the cloning breakthrough go to:  
[go.nature.com/tejh2](http://go.nature.com/tejh2)



► Lovell-Badge says that the mistakes still should have been caught, by both the authors and the reviewers and the journal, particularly one staffed by a team of professional editors. "It's the author's problem but if it's a journal like *Cell* or *Nature*, then I think, the journal, they have to take some responsibility to make sure there are no errors," he says.

Veronique Kiermer, executive editor at Nature Publishing Group, says that "this type of image manipulation is not something we expect referees to pick up. It's not really their role." She says that the *Nature* journals do "spot-checking" — on average, of two papers per issue. "But duplications between different figures, or between supplementary figures and the main text, are difficult to pick up," she says.

Mike Rossner, former executive director of the Rockefeller University Press, implemented image screening as part of the press's review process in its journals, such as the *Journal of Cell Biology*, in 2002. He says that all journals should do the same. Rossner says that inappropriately manipulated images appeared in as many as 25% of articles checked, although only about 1% affected the interpretation of the data. "I cannot speculate on whether our image-screening process would have caught these particular issues [in Mitalipov's *Cell* paper], but the process does include a visual check for duplicated regions in a figure," he says.

Marcus defends the fast peer review: "It is a misrepresentation to equate slow peer review with thoroughness or rigor or to use timely peer review as a justification for sloppiness in manuscript preparation," she wrote in her statement.

Mitalipov admits that a rush to publish might have led to mistakes that he compares to "typos." But if he had to do it over, he says he wouldn't take any more time or do anything differently — other than to "make sure the errors weren't there."

"We had four cell lines in December. We had five months to put together data. Most images and most figures were ready months in advance. The project was done, completed, written," he says.

"If you have a paper that's that hot, it will go through quickly. Other projects will take more time," he says. "It's better than [it] sitting on someone's desk for 6 months."

Mitalipov is arranging to give about a dozen other groups access to the cells, which would allow them to validate the cell lines. "The first thing we want to do is have people confirm our results," he said. "We are not hiding these cell lines." ■

*Additional reporting by Ewen Callaway, Erika Check Hayden and Richard Van Noorden.*

## CANCER

# Immunotherapy's cancer remit widens

*Combination therapies hold great promise, but at what cost?*

BY HEIDI LEDFORD

Drugs that unleash the power of the immune system on cancers are generating considerable optimism in industry, but still Andrew Baum thinks analysts are selling them short. In a 22 May report, Baum, the London-based head of global health-care research at the investment bank Citi, forecasts that in ten years the drugs will be treating 60% of cancers and earning US\$35 billion a year.

Three elements contribute to his bullishness: the drugs are showing signs of wider effectiveness; many patients will take them for years; and the prices are stratospheric (see 'Stiff medicine').

One of the first such drugs to be approved, Yervoy (ipilimumab), costs about \$40,000 per month in the United States, and £15,000 (\$23,000) in the United Kingdom, where health-care officials negotiated a lower price. So far, expensive immunotherapies have been approved only for treating melanoma and prostate cancer. But this weekend, at the annual meeting of the American Society of Clinical Oncology in Chicago, Illinois, investigators will present promising results from trials that indicate that immunotherapies could soon have a role in treating cancers of the lung, kidney and stomach.

A new strategy will add to the costs: regimens that combine the drugs with radiation therapy, genetically targeted drugs and other immunotherapies. The hope is that these other treatments will enhance the ability of the immune system to recognize the tumour, either by further stimulating the immune system, or by

damaging the tumour so as to release antigens that the immune system recognizes.

Yervoy, made by Bristol-Myers Squibb, headquartered in New York, is remarkable because a three-month course can send cancer into remission for years (see [go.nature.com/k1e3m2](http://go.nature.com/k1e3m2)). However, only about one-quarter of patients with advanced melanoma responds to the drug, which spurs tumour-killing T cells into action by blocking an inhibitory signal. "When Yervoy was first approved in 2011, we all recognized that it was an important moment for the field," says Jedd Wolchok, an oncologist at the Memorial Sloan-Kettering Cancer Center in New York. "But it was also a call to do better."

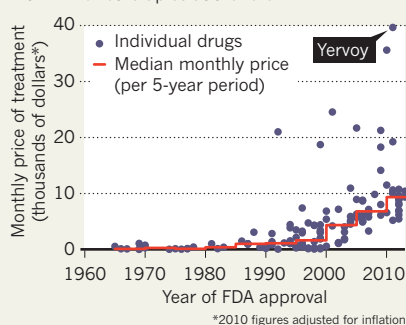
On 2 June, Wolchok will present data from a trial that combined Yervoy with an experimental drug called nivolumab, also made by Bristol-Myers Squibb. It releases a second brake on the immune system by stifling a different protein. The number of patients in the trial was small, but the results suggest that melanomas shrank in half of those who received the highest two doses tested.

Combination approaches could expand the number of cancers that respond to immunotherapy. But they could also accelerate the rise in the cost of cancer care. US spending on cancer drugs is rising by some 15% a year, twice as fast as health-care costs overall. "Cancer is a very complicated and expensive disease," says Scott Ramsey, a health-care economist at the Fred Hutchinson Cancer Research Center in Seattle, Washington. "But now it's turning into a chronic disease, and we're talking about years of maintenance therapy with drugs that cost \$10,000 a month." However, not all of the inflation is due to expensive new drugs, says Ira Klein, a medical director at insurance company Aetna, based in Hartford, Connecticut. The price of radiation therapy is rising by 25% per year, he notes, driven by new technologies that can cost \$100,000 or more for a full course of treatment.

Drug developers attribute the high cost of the drugs to the expense of research and development, which is compounded by the industry's high failure rate. Wolchok notes that drugs such as Yervoy have a small market, and speculates that prices may fall once their markets expand to other cancers. But Peter Bach, who studies health-care policy at Memorial Sloan-Kettering, is sceptical that industry will cut prices. "I have never seen that," he says. "I have only seen the opposite." ■

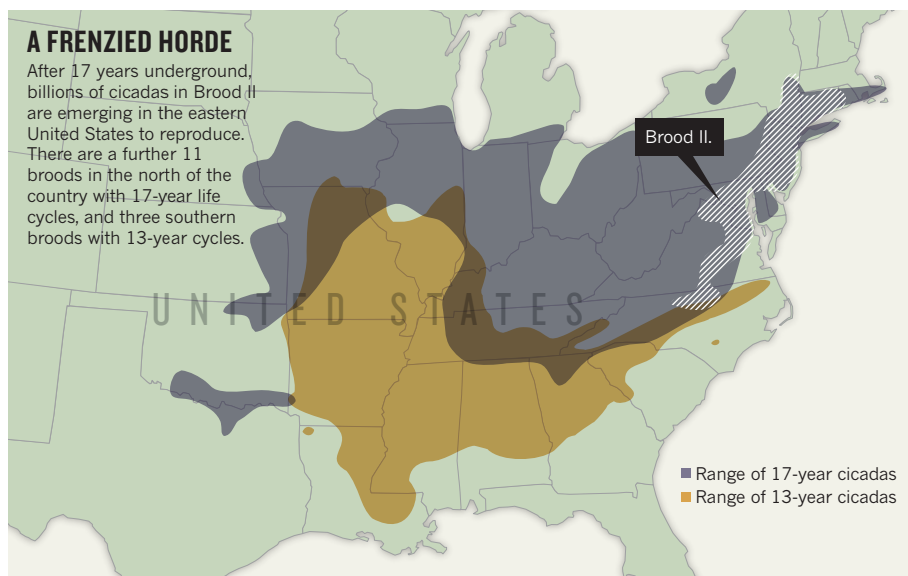
## STIFF MEDICINE

The cost of treating cancer is surging, with immunotherapies at the fore.



PETER BACH

SOURCE: G. KRITSKY



## ENTOMOLOGY

# Long-lived insects raise prime riddle

Mass orgy of 17-year cicadas sets US researchers buzzing.

BY RICHARD MONASTERSKY

**D**rivers who end up behind John Cooley this week will quickly lose their patience. Cruising around the eastern United States with his car window open, he slows down or stops every few hundred metres, cocks an ear and taps on a data-logger strapped into the passenger seat.

Since last week, Cooley, an evolutionary biologist at the University of Connecticut in Storrs, has been on the road mapping populations of periodical cicadas (*Magicicada*). These loud, red-eyed insects have spent the past 17 years maturing underground, only to emerge this month by the billions for a few weeks of singing and sex before they die. Like a handful of other cicada researchers on the prowl from North Carolina to New York,

Cooley knows that he has to work quickly. “Time is the real enemy here, for both the cicadas and the researchers,” he says. “If you miss this opportunity, you have a hole in your map and you have to wait for another 17 years.”

The insect genus with the longest known life cycles, *Magicicada* has confounded scientists for centuries. In 1665, the first volume of *Philosophical Transactions of the Royal Society* included a report from New England concerning

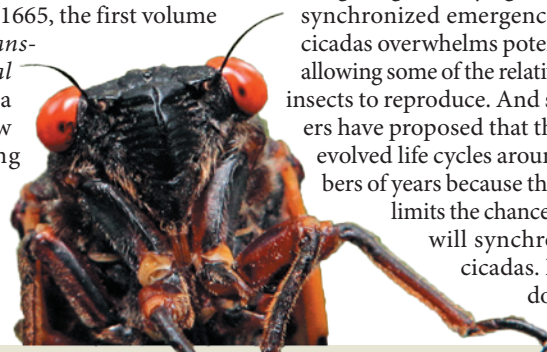
**Periodical cicadas emerge in broods of up to 350 insects per square metre.**

“swarms of strange insects, and the mischiefs done by them”. Charles Darwin also puzzled over them. Even now, entomologists are trying to understand how the insects’ peculiar life cycles evolved, how they count the years underground and how they synchronize their schedules. “They are one of the big ecological mysteries out there,” says Walt Koenig, a behavioural ecologist at the Cornell University Lab of Ornithology in Ithaca, New York.

They are also an entomological rarity. Of the thousands of cicada species known around the world, only the seven *Magicicada* species, which live in the eastern and central United States, have evolved such prolonged and synchronized life cycles (see ‘A frenzied horde’). At the southern end of their range, *Magicicada* populations have split into three mixed-species broods that emerge every 13 years. In the northern region, 12 broods follow 17-year cycles. This year’s crop belongs to Brood II, one of the larger 17-year groups and one that emerges in major metropolitan regions on the east coast.

Since their last outing in 1996, Brood II cicadas have grown through five larval stages underground, where they survived by sucking fluids from tree roots. With the warm weather this month, the nymphs have been crawling out of the ground before moulting for one last time and taking wing. Reaching densities of up to 350 individuals per square metre in woodlands, the cicadas can chorus at more than 95 decibels — loud enough to harm human hearing — as the males woo the females. After mating, the females cut slits into tree branches and lay their eggs there. By the time the eggs hatch, in six to ten weeks, the ground will be littered with the bodies of their parents. The new generations of nymphs will fall to the ground, burrow into the soil and remain there until 2030.

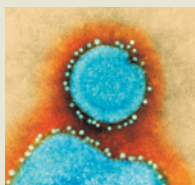
Biologists generally agree that the giant synchronized emergence of periodical cicadas overwhelms potential predators, allowing some of the relatively defenceless insects to reproduce. And some researchers have proposed that the cicadas have evolved life cycles around prime numbers of years because that arrangement limits the chances that predators will synchronize with the cicadas. But these ideas do not address ►



GERRY BROOME/AP PHOTO

  
**MORE ONLINE**

## TOP STORY



Emerging H7N9 flu easily transmits between ferrets [go.nature.com/pzwpst](http://go.nature.com/pzwpst)

## MORE NEWS

- Unusually strong deep-mantle earthquake strikes off coast of Russia [go.nature.com/kipwvj](http://go.nature.com/kipwvj)
- Graphene produces ultrashort laser pulses [go.nature.com/nl4ahd](http://go.nature.com/nl4ahd)
- Protein has key role in specialized itch neurons [go.nature.com/8pmbec](http://go.nature.com/8pmbec)

## PODCAST



New bird-like dinosaur; what sets Earth and Venus apart; and mosquitoes with no sense of smell [go.nature.com/mzu7it](http://go.nature.com/mzu7it)

CDC/SPL

► why the generations specifically last for 13 or 17 years.

Koenig suggests that the answer may involve interactions with birds. He and Andrew Liebhold of the US Forest Service in Morgantown, West Virginia, analysed 45 years of data from the North American Breeding Bird Survey (W. D. Koenig and A. M. Liebhold *Am. Nat.* **181**, 145–149; 2013) and found that bird populations tend to fall during the years in which periodical cicadas emerge. Birds feed on cicadas, so Koenig expected to find the opposite pattern. He proposes that the masses of cicadas trigger long-term changes in the forest that end up causing bird populations to crash after 13 or 17 years. The mechanism remains a mystery, but Koenig notes that one factor could be the flood of dead cicadas, whose bodies are 10% nitrogen. The die-off sends a pulse of fertilizer into the forest that temporarily enhances plant growth but could later lead to unfavourable conditions for birds. “It’s a pretty weird hypothesis,” he admits.

To synchronize their emergence, the nymphs must somehow keep track of how long they have been underground. Gene Kritsky, an entomologist at the College of Mount St. Joseph in Cincinnati, Ohio, says that nymphs seem to count the number of times that trees set their leaves in the spring; in 2007, some Brood XIV cicadas emerged a year early, following a strong winter thaw during which trees produced leaves, then dropped them and grew new ones in the subsequent spring. But no one knows how cicadas ‘remember’ the number of years since they last emerged.

Researchers are making more progress in



A cicada moults for the last time before taking wing.

probing the biological mechanisms that allow cicadas to switch their life cycles. In an analysis of DNA markers published this year (T. Sota *et al. Proc. Natl Acad. Sci. USA* **110**, 6919–6924; 2013), a team including Cooley developed an evolutionary tree for *Magicicada* and found that the major species groups had repeatedly split into 13-year and 17-year cohorts. The researchers suggest that those splits are explained by a common genetic mechanism across the species.

Chris Simon, a co-author and evolutionary biologist at the University of Connecticut,

plans to follow up those results with several genetic studies, including sequencing the RNA transcripts of genes that are active at different stages in the cicada life cycle. She is particularly interested in probing the occasional tendency of periodical cicadas to emerge 4 years early or late. These ‘stragglers’ are easy prey so do not usually survive, but Simon and others suggest that the timing mistakes might have given rise to new broods in the past. “It’s a way of having instant speciation,” she says. “This ability to jump through time is something that has not been seen before in other organisms.”

One example of this time travel might be happening right now in north-central Cincinnati, at least 500 kilometres from Brood II’s range. Kritsky documented thousands of cicadas appearing last week in a spot where he saw stragglers in 2000, four years before the city was inundated with the expected 17-year cicadas of Brood X.

The arrival of cicadas in the same place this year might mean that an environmental change such as global warming is causing them to emerge early, or that a genetic factor has caused some members of the 17-year Brood X to switch to a 13-year life cycle, says Kritsky. He must wait another 4 years to see whether some of the stragglers from 2000 have reverted back to their 17-year schedule. The main pulse of Brood X will arrive in 2021.

By then, Kritsky will be 68 years old. The long generations of the periodical cicadas makes studying them difficult, he says. “You would think we’d have a lot of answers but we don’t. Very few researchers have seen five generations.” ■

## GENOMICS

# Tepid showing for genomics X prize

*Challenge may be too hard and commercially unnecessary.*

BY ERIKA CHECK HAYDEN

It was never meant to be a piece of cake — but neither was it meant to be a flop. Yet as the 31 May registration deadline looms for the Archon Genomics X Prize — a challenge to sequence 100 complete human genomes in 30 days at unparalleled accuracy and low cost — only two teams have entered.

The lacklustre showing is a testament to both the difficulty of the challenge and the maturation of the DNA-sequencing industry in the seven years since the prize was first conceived,

genetics and innovation researchers say.

“The business has become bigger than the prize,” says Jonathan Rothberg, founder of the sequencing company Ion Torrent in Guilford, Connecticut, which was acquired in 2010 by Life Technologies in Carlsbad, California — which was, in turn, recently snapped up for US\$13.6 billion by Thermo Fisher in Waltham, Massachusetts. Ion Torrent plans to compete, but other firms have apparently decided that they have little to gain.

Yet the goal of the prize — to drive down the cost of sequencing while improving

its quality — matters just as much as it did in October 2006, when the X Prize Foundation, based in Playa Vista, California, first announced the challenge, experts say. Although sequencing costs have fallen drastically (see ‘Plummeting costs’), that decline has plateaued recently.

The original rules called for teams to sequence 100 genomes in 10 days for less than \$10,000 per genome. After none of the original eight competitors could meet the 10-day timeframe, the foundation spent two years revamping the challenge. The reconceived prize, launched in October 2011, extended the time to 30 days, tightened the cost to \$1,000 per genome and specified that 100 genomes from centenarians, who may harbour life-extending genetic variants, must be sequenced (see L. Kedes and G. Campamy *Nature Genet.* **43**, 1055–1058; 2011).

The new challenge aims at what the X Prize Foundation calls a “medical grade genome” — a sequence of all the nuclear DNA to 98% completeness and high accuracy, allowing only one error per million bases. To win the \$10-million prize pot, teams must also find



## OTHER CHALLENGES

## Popular bioinformatics

The X prize isn't the last word on genomics competitions. Cheaper to solve than technology-based contests, bioinformatics challenges have proved hugely popular.

The Critical Assessment of Genome Interpretation experiment has seen the number of entrants rise every year since it was first held in 2010 (see [go.nature.com/dfc1t1](http://go.nature.com/dfc1t1)). And a US\$1-million purse offered by the US Defense Threat Reduction Agency has spurred thousands of researchers to try to identify individual organisms from mixed pools of sequenced DNA.

Thousands of participants also competed

in a bioinformatics challenge in April to find hidden sequences in DNA data sets (run by the journal *Genome Biology* and curated by Michael Schatz of Cold Spring Harbor Laboratory in New York and James Taylor of Emory University in Atlanta, Georgia). It was organized to commemorate the sixtieth anniversary of the original research papers describing the structure of DNA.

Now Grant Campany, senior director for the Archon Genomics X Prize, says that the X Prize Foundation itself is considering laying out another challenge aimed at genome interpretation. **E.C.H.**

DNA insertions, deletions and rearrangements within genes and determine which parent each one came from.

Hitting all these goals in one go is hugely challenging. Market leader Illumina in San Diego, California, boasts a rate of false-positives (inaccurately flagging a DNA base as a variant from normal) of 0.25% and a rate of false negatives (missing a real variant) of 7.4%, well above the error rates allowed by the X-prize requirements. It is hard to do better because current technologies sequence the genome in short stretches that then have to be reassembled, introducing errors. New ways to sequence longer segments in one go are commercially available (see *Nature* **494**, 290–291; 2013), but they are slow and expensive. “To date, none of them would win the X prize at this scale,” says quantitative biologist Michael Schatz at Cold Spring Harbor Laboratory in New York.

Still, why have so few teams even deigned to try? Meeting the challenge would cost much more than the prize purse, but that has also been true of past contests that attracted dozens of entrants, such as the Ansari X Prize, which required teams to send passengers into space, and the US government's Defense Advanced

Research Projects Agency Grand Challenges, one of which catalysed the development of successful self-driving cars.

Part of the answer is that a genomics prize, unlike a rocket launch, isn't easy to explain to the public. As a result it does not have the same publicity value, says Luciano Kay, a researcher at the Center for Nanotechnology in Society at the University of California, Santa Barbara. A competition for a self-driving car that can go 10 kilometres is more attractive than manipulation of matter or genes at tiny scales to accomplish a very scientific or technical feat, Kay says.

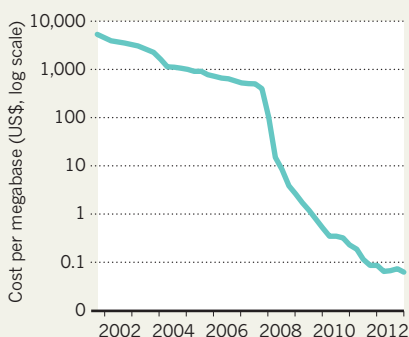
And the goal of the genomics X prize — to sequence whole genomes to medical grade rapidly and cheaply — may not be a top commercial priority at present. The business of genomics is already booming on the basis of less complete sequences, and Rothberg points out that scientists can only interpret the small fraction of the genome that codes for proteins (the ‘exome’). It is unclear what would be gained from an accurate sequence of the rest.

Still, whoever wins the prize earns the right to boast, which explains why Ion Torrent decided to compete. It also explains why Illumina decided not to: failure would only dent its reputation, muses Timothy Harris, who develops applied-science tools at the Howard Hughes Medical Institute's Janelia Farms campus in Ashburn, Virginia. The other entrant is a team led by George Church at the Wyss Institute at Harvard in Boston, Massachusetts.

Grant Campany, senior director of the genomics X prize, hopes that other teams will step up to compete before the contest gets under way in September. Even if they don't, scientists predict that its goals will be achieved within the next few years, whether through the prize or not. “If you could deliver that kind of performance you would have the commercial advantage by a large margin over anyone else,” Harris says. “That commercial advantage is worth way more than the X prize.” ■ **SEE EDITORIAL P. 535**

## PLUMMETING COSTS

Advances in sequencing technologies have driven a sharp drop in price.



SOURCE: NHGRI

# Japan aims high for growth

*Innovation in science is at the heart of government plans to boost the economy.*

BY DAVID CYRANOSKI

The Japanese government is working on a plan to revitalize its science workforce by boosting opportunities for female scientists, attracting top talent from abroad and increasing the commercialization of research. So what else is new? Over the past decade, successive administrations have had similar goals, but little progress has been made. This time, analysts and scientists think that things might be different.

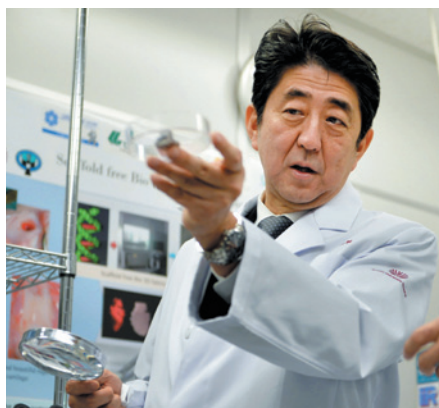
Prime Minister Shinzo Abe is riding high since his Liberal Democratic Party swept back to power in December. He has stimulated the economy, ending 15 years of deflation, jump-started the stock market and weakened the yen to spur exports. His reform platform includes a new growth strategy, and central to that strategy is innovation in science and technology.

Abe's cabinet has already committed hundreds of billions of yen to space, physics and stem-cell research, in a stimulus package announced in January. But more aggressive measures are yet to come. On 17 May, the Council for Science and Technology Policy — the nation's leading science body, which Abe chairs — released the first draft of a Comprehensive Science, Technology and Innovation Strategy, calling for even more fundamental changes.

To be finalized by the end of this year, the strategy includes eye-catching targets for bolstering the country's intellectual base. It calls for 30% of new recruits at research institutes and universities to be female by 2016, up from 21.2% in 2010. And faced with fewer university graduates and a shrinking birth rate, Japan is looking to open itself up to foreigners. Under the plan, international researchers would occupy 20% of staff positions at leading research organizations by 2020, and 30% by 2030. That would be an ambitious jump from the 2010 figure of 3.9%.

To increase technology transfer, the plan sets a goal of doubling the number of collaborations worth more than 10 million yen (US\$98,000) between university and industry by 2030. And it calls for the numbers of foreign patent applications and collaborations lasting more than 3 years to similarly double.

Japan needs to make up ground. It has lost the competitive edge it once had over China, South Korea and other Asian rivals in industries such as microelectronics and pharmaceuticals. "Over the past decade, Japan has been stagnant in terms of innovation," says Yuko Ito, head of the Science, Technology and Innovation Policy division at the Tokyo-based



Japan's Prime Minister Shinzo Abe visits a lab in Fukuoka.

National Institute of Science and Technology Policy.

Abe failed to reverse those trends in 2006, during his first term as prime minister, with a largely ineffectual 'Innovation 25' programme. Nonetheless, some see reasons for optimism, not only about economic success but also for a more innovative workforce. The strategy this time "is emphasizing the need for cultivating human resources, especially women," says Ito.

Hisako Ohtsubo, a molecular biologist at Nihon University near Funabashi who researches gender equality in science, is also cautiously optimistic. She says that it was a pleasant surprise to hear Abe — a conservative who, for example, opposes passage of the imperial throne to the female blood line — repeatedly mention the importance of increasing women's role in the business world. Most startling for Ohtsubo was Abe's statement that childcare across Japan should be expanded, something many analysts have said is crucial for women's ability to have careers in research and other sectors. "He's a different phenotype," she says. "Before he would never have talked about such things."

Still, she is waiting to see whether Abe will follow up his words with the necessary investment in programmes to give extra grants to women and encourage companies and universities to hire women to senior researcher positions, and not just as regular staff. "That's the only way we will be able to overcome the deeply embedded stereotypes in the system," Ohtsubo says.

Piero Carninci, a genomics expert based in Yokohama at the RIKEN Center for Life Science Technologies with nearly two decades' experience in Japan, has advice for Abe if he is serious about courting foreigners. As head of RIKEN's new Division of Genomic Technologies, he

is the first foreigner to hold a division director position. The top-down management style of mentors, communication problems in the laboratory and the Japanese lifestyle can all be daunting for people from abroad, he says.

Carninci's remedy includes giving foreign researchers academic independence and reasonable start-up budgets, as well as assistance in overcoming language problems. Also key: offering an equal opportunity with Japanese nationals to climb the career ladder.

Atsushi Sunami, a science-policy expert at the National Graduate Institute for Policy Studies in Tokyo, says that efforts to hire overseas talent may also benefit from one of the themes of the government's growth strategy: deregulation of markets and quasi-governmental organizations, including universities.

Decisions on the main budget items still have to be made, Sunami says. A proposal in the comprehensive strategy would give an annual budget — some 50 billion yen, according to Sunami — to the Council for Science and Technology Policy to fund science. Currently, the council is only an advisory body. But until the budget decisions are made, Sunami says, the strategy is "just a piece of paper".

Koichi Sumikura of the National Institute of Science and Technology Policy thinks that the Fukushima nuclear disaster, and the ensuing energy shortages, should help to keep innovation on the agenda. "The disaster is what really makes things different this time," says Sumikura. "Before, politicians talked about innovation without indicating any real direction. Now there is a clear need to recover from the disaster and build a stronger society." ■ SEE EDITORIAL P.535

## CORRECTIONS

The world map in 'GM crops: A story in numbers' (*Nature* **497**, 22–23; 2013) wrongly highlighted Slovenia instead of Slovakia as one of Europe's GM-crop-producing nations. The table in the News story 'Log-jam in agency confirmations' (*Nature* **497**, 418–419; 2013) omitted Christine Todd Whitman's full name. In the News story 'Graphene knock-offs probe ultrafast electronics' (*Nature* **497**, 422–423; 2013), reference 2 should have cited volume 438, not 483. And the News Feature 'Outward bound' (*Nature* **497**, 424–427; 2013) wrongly stated that Voyager 1 and Voyager 2 were renamed en route to Jupiter. In fact, they were renamed before launch.

THE ASAHI SHIMBUN VIA GETTY





# Method man

Karl Deisseroth is leaving his mark on brain science one technique at a time.

BY KERRI SMITH

When Karl Deisseroth moved into his first lab in 2004, he found himself replacing a high-profile tenant: Nobel-prizewinning physicist Steven Chu. “His name was still on the door when I moved in,” says Deisseroth, a neuroscientist, of the basement space at Stanford University in California. The legacy has had its benefits. When chemistry student Feng Zhang dropped by looking for Chu, Deisseroth convinced him to stick around. “I don’t think he knew who I was. But he got interested enough.”

Deisseroth is now a major name in science himself. He is associated with two blockbuster techniques that allow researchers to show how intricate circuits in the brain create patterns of behaviour. The development of the methods, he says, came from a desire to understand mechanisms that give rise to psychiatric disease — and

from the paucity of techniques to do so. “It was extremely clear that for fundamental advances in these domains I would have to spend time developing new tools,” says Deisseroth.

His measured tone and laid-back demeanour belie the frenzy that his lab’s techniques are generating in neuroscience. First came optogenetics<sup>1</sup>, which involves inserting light-sensitive proteins from algae into neurons, allowing researchers to switch the cells on and off with light. Deisseroth developed the method shortly after starting his lab, working with Zhang and Edward Boyden, a close collaborator at the time. Optogenetics has since been adopted by scientists around the world to explore everything from the functions of neuron subtypes to the circuits altered in depression or autism. Deisseroth has lost count of how many groups are using it. “We sent clones

RAMIN RAHIMIAN/GETTY FOR NATURE



to thousands of laboratories,” he says.

Now his lab is gearing up for another rush, after publishing a method called CLARITY<sup>2</sup>. The technique uses a chemical treatment to turn whole brains transparent, allowing researchers to examine the brain's structure in detail in three dimensions. It could help researchers to connect the brain's function with its structure. “I think it'll be de rigueur for people to have a picture showing the pathways and how they sit in the brain,” says Richard Tsien, the neuroanatomist under whom Deisseroth earned his PhD at Stanford.

Deisseroth takes risks and encourages his team to do the same, often pulling together people with varying expertise and keeping abreast of many different fields. Christof Koch, chief scientific officer at the Allen Institute for Brain Science in Seattle, Washington, puts Deisseroth's success down to these “catholic interests”. But Deisseroth also has a laser-like focus on tool-building, fuelled, he says, by his work as a practising psychiatrist. And although he is excited about his lab's newest technique, Deisseroth plays it down, shrugging a broad shoulder: “I'm not very demonstrative,” he says.

## TAKING CONTROL

When Tsien, who now works at New York University's Langone Medical Center, first met Deisseroth, he was struck by his determination. “It was an unusual encounter — he came for an interview and told me rather shyly that the reason he came to Stanford was because he wanted to work with me,” Tsien remembers. It was 1993 and Tsien's lab was “crammed to the gills”, he says, but Deisseroth was persistent and persuasive. As part of a joint programme to earn a PhD and a medical doctorate, he began a project looking at calcium channels in neurons. In 2000, he followed that up with a multi-pronged appointment: a postdoctoral fellowship in the lab of Stanford neuroscientist Robert Malenka and — after toying with the idea of neurosurgery — a residency in psychiatry at Stanford Medical School.

“Everything changed when I did my psychiatry rotation,” says Deisseroth. “A person can be right in front of you who looks intact, not obviously injured, and yet their brain is constructing for them a completely different reality. At the same time I saw how deep the suffering was.”

Studying depression or anxiety in a dish of cells was inadequate, he reasoned, because only whole brains can give rise to the sophisticated functions — and disorders — that characterize human behaviour. And techniques for studying whole brains in humans and model organisms were often limited to simply watching them at work.

So Deisseroth began thinking about ways to examine and control intact systems. “I was having a lot of discussions with a lot of people,” he says. During his residency, Deisseroth met

Boyden, a PhD student with similarly ambitious aims. The two began talking about ways to manipulate individual neurons as a side project. “It was a very adventurous collaboration, full of exploration,” says Boyden.

One idea involved using light to control neuronal firing. Boyden and Deisseroth knew about light-sensitive channel proteins called opsins, which algae use for generating energy,

**“It was a little hard for people in the field to understand what he was trying to do.”**

among other functions. Several groups — including that of Tsien's brother Roger at the University of California, San Diego — were trying to insert these proteins into neuron cell membranes. The project needed “the wherewithal to spend the money and find the graduate students”, says Deisseroth. In 2004, having secured his own lab, he could do just that.

By July of that year, Deisseroth had managed to coax neurons into expressing opsins in their cellular membranes. Using cells borrowed from Malenka and recording equipment from Richard Tsien, he, Zhang and Boyden hunkered down to see if the technique would fly. “It worked on pretty much the first try,” Boyden recalls. “It was like riding a wave of serendipity.”

In the following year, Deisseroth finally secured funding from the US National Institutes of Health (NIH) to further the work. He had faced multiple rejections. Projects based on technologies have a harder time attracting support than hypothesis-driven projects, says Thomas Insel, director of the US National Institute for Mental Health in Bethesda, Maryland, the source of the initial federal funding for Deisseroth's optogenetics research. “It was a little hard for people in the field to understand what he was trying to do,” says Insel.

But neuroscientists saw the potential in 2005, when Deisseroth's group published its first big paper showing that the technique worked in a dish<sup>1</sup>. Researchers could now stimulate a specific type of neuron and see how that affected the cell's — or even an entire animal's — behaviour. Requests for the technology were pouring in.

The technique has since been used for everything from studying the development of neural stem cells<sup>3</sup> to prompting mice to recall fearful memories<sup>4</sup> (see *Nature* 465, 26–28; 2010). Deisseroth and his team, with their focus on psychiatric conditions, have used rodent models to explore the network of brain areas that gives

rise to anxiety and have shown that one ‘hub’ controls diverse symptoms such as an elevated

breathing rate and feelings of panic and discomfort<sup>5</sup>. They have switched on and off the activity of mouse neurons that use the neurotransmitter dopamine, to show how they contribute to the symptoms of depression<sup>6</sup>. The team has even used optogenetics to prevent cocaine-addicted rats from seeking out the drug<sup>7</sup>. The work, says Deisseroth, could help scientists to design anti-anxiety drugs that are less addictive than current treatments — affecting anxiety pathways but leaving reward circuitry untouched.

In 2006, Boyden started his own lab at the Massachusetts Institute of Technology in Cambridge. Rumours emerged that his relationship with Deisseroth was souring. Neither is keen to discuss the details, but in 2007, the two groups published separate papers on halorhodopsin, an ‘off-switch’ opsin that they had previously studied together<sup>8,9</sup>. “We didn't know Ed was working on that until we sent our paper to *Nature*,” says Zhang. “Well you know, in any competitive scenario people are interested in being first,” says Boyden.

The experience might have left its mark on Deisseroth: he is careful to attribute credit precisely to his team members and trainees, particularly with respect to CLARITY. “Karl is quite sensitive about who came up with the idea,” says Kwanghun Chung, first author of the CLARITY paper<sup>2</sup>. Creating see-through brains was Deisseroth's dream, but the chemical know-how came from Chung. They describe the finding as a co-discovery.

## SEEING CLEARLY

Standard techniques for examining fine-scale brain structure involve slicing the brain into tissue-thin segments, analysing them under a microscope, and then — laboriously and often imprecisely — stacking the images back together. To bypass that process, Deisseroth had been looking for a means of chemically treating the brain to make it transparent. Some components of brain cells — especially lipids — are notoriously opaque to microscopes, so the idea was to get rid of them, while preserving the neuronal structure. It was a very different approach from optogenetics, and recruiting Chung — a chemical engineer by background — displayed Deisseroth's willingness to take risks. “I remember people wondering what I was doing, what sort of plan there was in hiring this person and the direction the lab was going, because it was such a marked shift,” he says.

At first, Deisseroth wanted to immobilize neurons by filling them with a suitable material and removing the surrounding tissue. It turned out to be very difficult to build a stable structure from inside the neurons, but it was easier to build one around them.

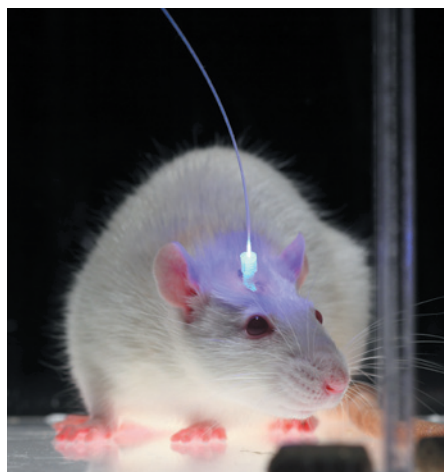
The team tried a few scaffolding materials, including keratin and a cellulose-like structure, before settling on a hydrogel — a gel made mainly of water and already widely used in biological studies. The researchers found that a

## ➔ NATURE.COM

For interviews with Deisseroth and a video of CLARITY, see: [go.nature.com/zsxdld](http://go.nature.com/zsxdld)

## A light touch

Karl Deisseroth's lab has produced two blockbuster techniques that allow researchers to visualize and manipulate the brain.



### Optogenetics

- 1 Viruses are used to ferry genes encoding light-sensitive receptors (opsins) into specific neurons.
- 2 Animals are fitted with an 'optrode', a fibre-optic cable with an electrode.
- 3 Light beamed down the optrode will open or close ion channels while the electrode records neuronal firing and researchers record behaviours.

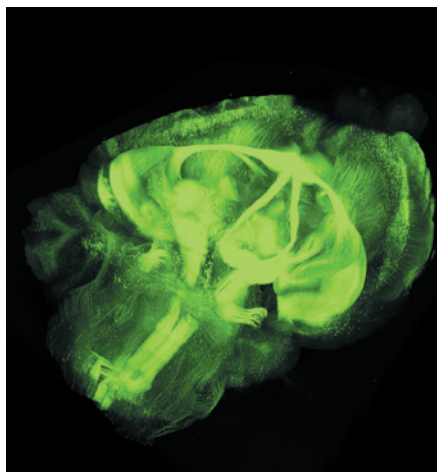
hydrogel scaffold could bind components of neurons in place, including proteins, neurochemicals and even DNA and RNA (see 'A light touch'). With the structure locked down, a detergent called SDS could wash away lipid membranes, leaving the tissue transparent (see *Nature* **496**, 151; 2013).

Within a week of publication, Deisseroth had already received several dozen requests for information. "We've been very much... I'm trying to find a verb that doesn't make it seem like a problem," he says. He settles on describing the e-mails that the lab received as an "outpouring".

CLARITY stunned people in the same way that optogenetics did, says Insel. "Usually when someone has at an early stage of their career made an important advance, they tend to rest on those laurels. Karl quite quickly decided to do the next big thing."

Other researchers will tailor CLARITY to their own projects. David Van Essen, a neuroscientist at Washington University in St. Louis, Missouri, is interested in testing CLARITY in the brain's white matter, the bundles of neuronal projections called axons that carry nerve impulses. This will help his group study patterns of connectivity that link brain regions.

Using CLARITY with techniques such as optogenetics, or using it to analyse brains after behavioural studies, will be a powerful way to extract information about how brain networks



### CLARITY

- 1 Proteins and other molecules in the brain are hybridized to a hydrogel matrix, fixing them in place.
- 2 Densely packed, light-blocking lipids are washed away using detergents.
- 3 The clarified brain can be stained and washed multiple times, allowing scientists to observe different molecules and cell types.

function, says Eve Marder, a neuroscientist at Brandeis University in Waltham, Massachusetts. But analysing large networks in detail is challenging. Marder usually works with simple circuits of 30 or so neurons, and says that even those have a scarily exponential number of permutations. "My only personal hope is that people who venture into doing circuit work in larger brains pay attention to the lessons learned and mistakes that we made along the way," she says.

CLARITY is already shedding light on clinical disorders. In the paper describing the technique<sup>2</sup>, the researchers analysed brain tissue from a seven-year-old boy who had autism spectrum disorder. They found that neurons in his cortex had joined together in ladder-like patterns, rather than the branches seen in typical brains. Animal models of autism-like conditions had hinted at this difference, but CLARITY made it possible to look for the irregularity in human brain samples.

#### RISK AND REWARD

Deisseroth's office and much of his lab are still in the same basement that he moved into in 2004, although the team — now including 35 people — has branched out into lab space elsewhere. The labs are strewn with gadgetry, like the bedroom of a technology-obsessed teenager. There is no natural light in the basement, but Deisseroth likes it; being at the bottom of

the building keeps vibrations to a minimum, which is necessary when using the sophisticated microscopes essential for his work.

The current team includes computational neuroscientists, medics, chemists and engineers. And a US\$22.5-million Transformative Research Award from the NIH, granted last year, allows Deisseroth to keep encouraging them to take chances. "I have everyone have a hand in a high-risk and a low-risk project," he says, "like a hedged portfolio."

One of the latest high-risk projects uses light-field microscopy, which records images from a variety of perspectives simultaneously, to image the brain. Elsewhere in the lab, Chung and collaborator Viviana Gradinaru are still trying to build hydrogel structures inside neurons in the hope of preserving networks of specific cell types — freezing in place all the neurons of a certain type, for example, or all the cells that express a particular gene. The team continues to improve optogenetics, too, developing new opsins and perfecting methods for delivering light to the brain.

Deisseroth has already added space to the laboratory to cater for all the visiting scientists coming to learn how to use optogenetics and CLARITY. There were more than 200 in 2010–12. The group has "done yeoman's work in making this relatively straightforward to set up", says Van Essen. "They're not going to hold hostage a methodology that people really can set up in their own lab." The tools are freely available; Stanford has filed patents on some aspects, says Deisseroth, but that it is to ensure unfettered access to the technology. He does not make money from the patents.

Deisseroth struggles with his decision to spend more time on research than with the patients who inspire it. He still runs a weekly psychiatry clinic, but has to balance this with a packed travel schedule and mentoring duties. He carries on, however, hopeful that his studies will ultimately benefit those that medicine cannot. "We can help patients right now with many psychiatric symptoms," he says. "But for others we can't. So that really helps me in mentally freeing up the time needed for research."

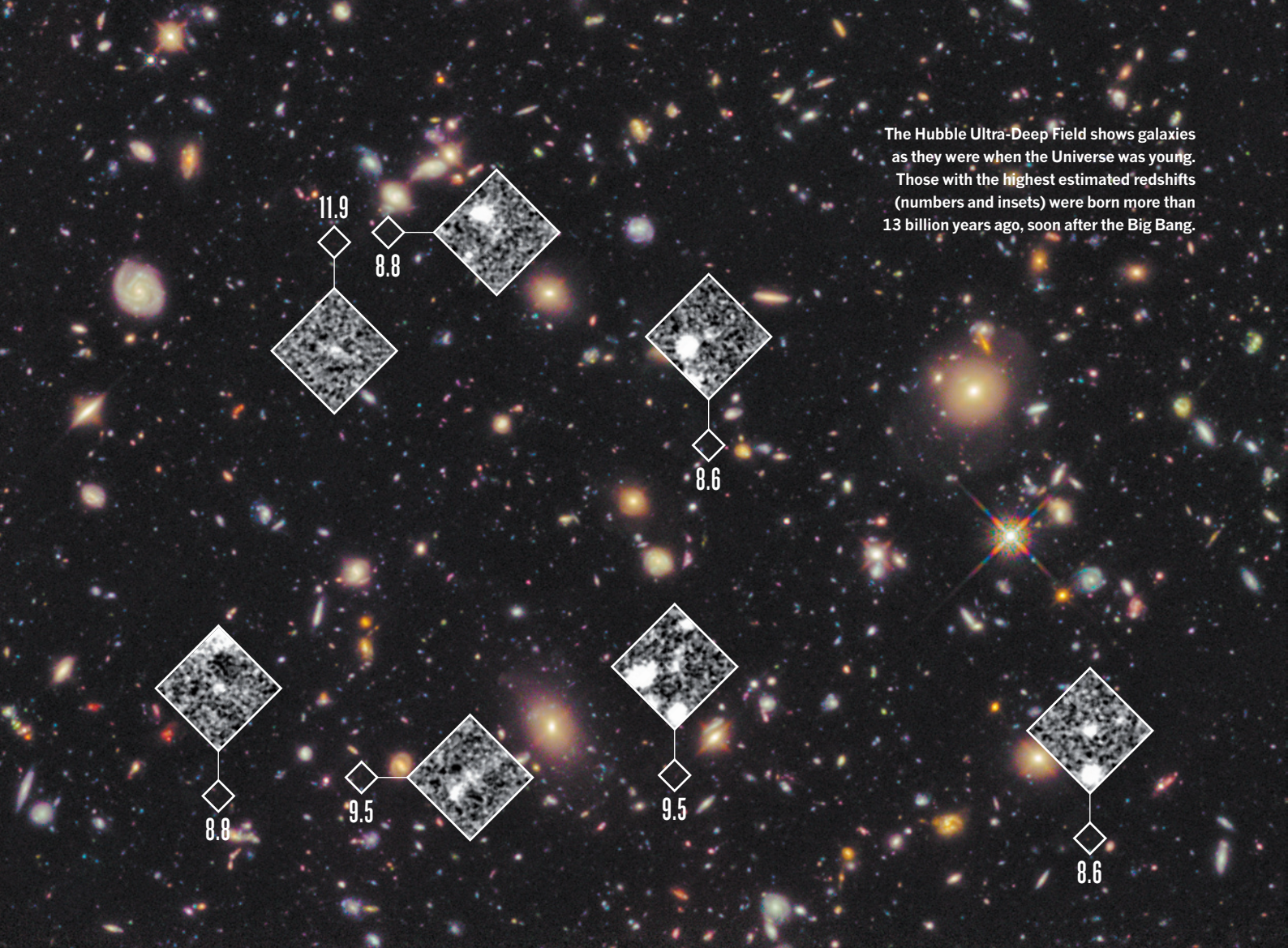
That desire to help is buoyed by simple, boundless curiosity. "I think all scientists have a little bit in them," he says. "That wonder and curiosity about the natural world — not only how it is, but how it could be." ■

**Kerri Smith** is senior audio editor for *Nature* in London.

1. Boyden, E. S., Zhang, F., Bamberg, E., Nagel, G. & Deisseroth, K. *Nature Neurosci.* **8**, 1263–1268 (2005).
2. Chung, K. *et al. Nature* **497**, 332–337 (2013).
3. Strohm, A. *et al. Stem Cells* **29**, 78–88 (2011).
4. Liu, X. *et al. Nature* **484**, 381–385 (2012).
5. Kim, S.-Y. *et al. Nature* **496**, 219–223 (2013).
6. Chaudhury, D. *et al. Nature* **493**, 532–536 (2013).
7. Stefanik, M. T. *et al. Addiction Biology* **18**, 50–53 (2013).
8. Han, X. & Boyden, E. S. *PLoS ONE* **2**, e299 (2007).
9. Zhang, F. *et al. Nature* **446**, 633–639 (2007).

LEFT: V. GRADINARU/J. CARNETT/M. MOGRIG/K. DEISSEROTH/STANFORD UNIV.; RIGHT: K. CHUNG/K. DEISSEROTH/STANFORD UNIV.





The Hubble Ultra-Deep Field shows galaxies as they were when the Universe was young. Those with the highest estimated redshifts (numbers and insets) were born more than 13 billion years ago, soon after the Big Bang.

# COSMIC DAWN

*The Hubble Space Telescope is giving astronomers a glimpse of the Universe's first, tumultuous era of galaxy formation.*

**F**or one sleepless week in early September 2009, Garth Illingworth and his team had the early Universe all to themselves. At NASA's request, Illingworth, Rychard Bouwens and Pascal Oesch had just spent the previous week staring into their computer screens at the University of California, Santa Cruz, scanning through hundreds of black-and-white portraits of faint galaxies recorded in a multi-day time exposure by a newly installed infrared camera on the Hubble Space Telescope. NASA simply wanted the three astronomers to preview the images and make sure that the camera was working correctly, before the agency released the data more widely.

But Illingworth, Bouwens and Oesch were hoping that they would

BY RON COWEN

find more — that at least some of those smudges of light would prove to be among the first galaxies to form in the Universe, less than 1 billion years after the Big Bang. Even a faint glimpse of such objects could provide fresh insights into some of the biggest questions in cosmology, ranging from the nature of the first stars to the tumultuous beginnings of galaxy formation.

That week, the astronomers began to focus on two dozen tiny candidate images — each so dim and grainy that they might easily be noise in the camera's digital sensors. But as their analysis proceeded, it became clear that these patches of light had the right colour, appearing only in the camera's reddest filters — exactly what would be expected of newborn galaxies seen at a very great distance and very high redshift. And



when the three colleagues started digitally adding together exposures of each candidate, says Illingworth, “suddenly there they were” — fuzzy, but undeniable images of galaxies. “That week in September was one of the most exciting times of my career!”

By the week's end, he, Bouwens and Oesch had posted two draft papers to the arXiv preprint server<sup>1,2</sup>, detailing their first-ever collection of more than 20 galaxies from the age of galaxy formation, some 13 billion years ago, when the cosmos was only 600 million to 800 million years old. Since then, other researchers have made further observations of the same small patch of sky, known as the Hubble Ultra-Deep Field (HUDF), and four other larger regions. They have expanded that initial roster to some 1,400 young galaxies, from the same era.

The data from this growing catalogue are already hinting at a still-unseen time — an infant Universe thronged with countless small galaxies and lit by primordial stars so massive that they burned out and blew up in a cosmic eye-blink. And a new generation of instruments promises to bring that era into clear view. They include the Atacama Large Millimeter/submillimeter Array (ALMA) of radio telescopes in Chile, which is already beginning such observations; and Hubble's successor, the infrared James Webb Space Telescope (JWST), which is set for launch in late 2018.

It's a heady time for early-Universe astronomers, says cosmologist Avi Loeb of Harvard University in Cambridge, Massachusetts. “We're looking at our origins,” he says. “The first galaxies were the building blocks of the Milky Way, and the desire to understand them is a search for our roots.”

## DEEP BACKGROUND

Over the past few decades, observers have developed a general storyline describing how galaxies formed (see ‘Dawn's early light’). Astronomers know, for example, that the raw material was a hot, ionized plasma of hydrogen and helium that emerged from the Big Bang, then rapidly cooled as the Universe expanded. Once its temperature had fallen far enough, about 370,000 years after the Big Bang, protons and electrons combined to make neutral atoms and created a light-absorbing haze that plunged the Universe into a cosmic ‘dark ages’.

Astronomers also know that this cosmic haze was almost perfectly uniform at the start — but immediately began to clump together as gravity began to magnify slight fluctuations in the material's density. And they are reasonably sure that, after several hundred million years, the densest of the growing clumps began to form the first stars, which ignited by thermonuclear fusion and reionized the neutral gas that remained. The veil of gas became a transparent plasma again, bringing the cosmic dark ages to a spectacular end (see *Nature* **490**, 24–27; 2012).

But from this point onwards, very little is certain. The formation of succeeding generations of stars and galaxies was a swirling chaos of heating and cooling gas clouds, detonating supernovae, black-hole accretion and fierce stellar winds strong enough to eject matter from small galaxies — a process far too messy and complex to understand without extensive observations.

Such observations are a major goal of the HUDF project, which aims to gather enough images of distant galaxies to discern patterns in their sizes, shapes and colours. Located south of Orion in the constellation Fornax, and measuring just one-tenth of the diameter of the full Moon as seen from the ground, the HUDF is an otherwise typical patch of dark sky that happens to be relatively devoid of foreground stars

and galaxies. But just as astronomers expected, an 11.3-day time exposure of the field taken by Hubble in late 2003 and early 2004 revealed that it was, in fact, filled with a multitude of faraway galaxies seen as they were billions of years ago.

In August and September 2009, the field was

re-examined in an additional two-day exposure taken by Hubble's Wide Field Camera 3 (WFC3): an instrument installed by astronauts the previous May that is exquisitely sensitive at infrared wavelengths — exactly where visible and ultraviolet light from the farthest galaxies is expected to end up after being redshifted by the cosmic expansion.

These were the images that Illingworth, Bouwens and Oesch saw. Knowing that the WFC3 could detect distant galaxies about 30 times fainter than its predecessor could, or about 4 billion times fainter than anything visible to the human eye, the astronomers initially thought that they might have caught one of the very first generations of galaxies in the act of being born. When they estimated the objects' distance and composition by examining their colours in three different filters — the faint smudges were far too dim for Hubble to get a spectrum — the team found that they were relatively blue, exactly as expected of extremely young galaxies glimpsed in their first frenzy of star formation.

But this conclusion was far from ironclad. Testing the idea was a prime motivation for a team of astronomers led by Richard Ellis at the California Institute of Technology in Pasadena. In 2012 they re-examined a small part of the centre of the HUDF, this time with an additional colour filter and a time exposure totalling about 23 days.

These newer observations, which Ellis's team reported in January this year at a meeting of the American Astronomical Society in Long Beach, California<sup>3–5</sup>, reveal that the galaxies are in fact redder, and therefore contain older stars, than initially calculated. The very youngest galaxies that Hubble has identified, imaged as they appeared 560 million to 780 million years after the Big Bang, contain stars that are 100 million to 200 million years old. So these galaxies had already been around for at least that long.

The new HUDF observations also reveal puzzling features of the tumultuous era of reionization, as explained at the January meeting by Brant Robertson of the University of Arizona in Tucson. This was the time when the first galaxies were growing bigger and more numerous, and when ultraviolet light from the first stars was becoming strong enough to ionize the veil of thick hydrogen gas that enveloped them. Other observations show that reionization began roughly 250 million years after the Big Bang, and that it was complete at a cosmic age of roughly 1 billion years — at which point starlight could stream freely into space and the cosmos was mostly transparent, just as we see it today.

But although the galaxies Hubble saw in the 2012 (and the 2009) HUDF observations were presumably the largest and brightest ones around all those billions of years ago, there simply were not enough of them to reionize the Universe. This means, according to Ellis, Robertson and their colleagues, that there must have been a large population of unseen small fry that did most of the work — a conclusion also reached by Illingworth and his team<sup>6</sup>.

“We now know there's a whole population of small galaxies at even earlier times” than Hubble's detectors can record, says Ellis — which leads to an exciting set of questions for the newer telescopes such as ALMA and JWST, including how these bodies formed and how they coalesced into the larger galaxies that came later.

Another set of questions relates to the very first generation of stars, which coalesced from almost pure helium and hydrogen forged in the Big Bang. Theory suggests that they were more than 100 times as massive as our Sun — far larger than any stars that form today. If so, then theory also suggests that these monsters were so short-lived that none of them would have survived in the galaxies that Hubble can see. Their extreme size would have caused these stars to destroy themselves in spectacular supernova explosions after only some 2 million years. But did they? And did their death throes delay the birth of the next generation of stars, by disrupting the thick interstellar gas clouds in which new stars were forming?

The HUDF data already suggest that the answer to the last question

“SUDDENLY THERE  
THEY WERE — FUZZY,  
BUT UNDENIABLE  
IMAGES OF GALAXIES.”

## ➔ NATURE.COM

To watch a simulation of galaxies forming see: [go.nature.com/tz1dpt](http://go.nature.com/tz1dpt)

## DAWN'S EARLY LIGHT

The Big Bang generated a haze of primordial hydrogen and helium. After about 370,000 years, cosmic expansion cooled it enough for electrons and protons to form neutral atoms ('recombination'). During the resulting 'dark ages', when the Universe had no stars to light it, gravity magnified slight fluctuations in density and caused the neutral gas to clump. The densest clumps eventually became the first stars and galaxies. Their ultraviolet light ionized the remaining neutral gas ('reionization'), allowed light to stream freely into space, and brought the dark ages to an end.

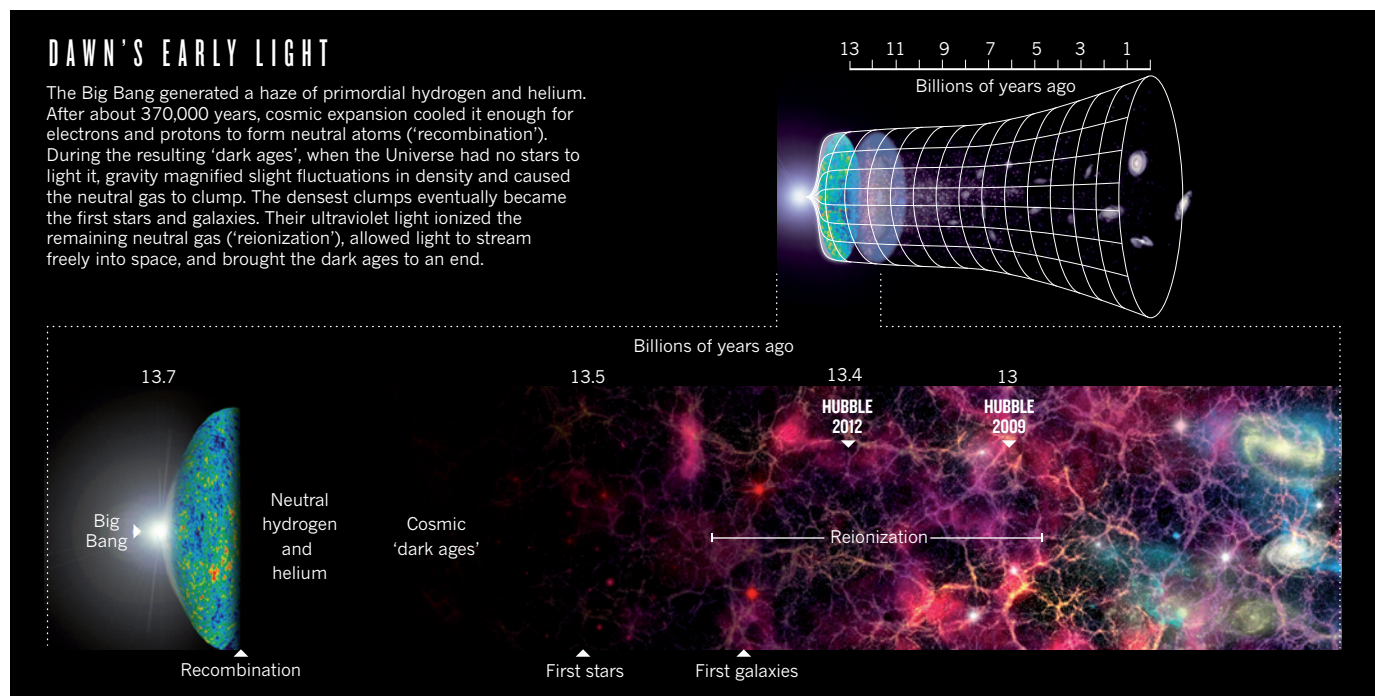


ILLUSTRATION: NIK SPENCER; SOURCES: NASA/WMAP SCIENCE TEAM; R. ELLIS (CALTECH)

is 'No', says theorist Volker Bromm at the University of Texas at Austin, who was not involved in the 2009 and 2012 HUDF studies. Because the colours of the galaxies seen in the ultra-deep field indicate that they had been forming stars for at least 100 million years already, this suggests that there was little or no lag between the death of the very first generation of stars and the birth of the second, he says. The generations may even have overlapped. But untangling exactly what happened will be a job for future telescopes.

## THE NEXT FRONTIER

In the meantime, Hubble astronomers haven't been idle. NASA is pursuing a tactic that could turn the observatory into a telescope as powerful as the JWST will be, along some limited fields of view.

To achieve this, astronomers are scanning the heavens to select six fields of view that, unlike the HUDF, each contains a high-mass cluster of foreground galaxies. As first predicted by Einstein, such clusters act as cosmic zoom lenses, gravitationally magnifying and brightening images of distant galaxies that lie directly behind them.

Hubble's visible-light and infrared cameras will take turns at looking through these lenses, which should reveal distant galaxies 10 to 50 times fainter than any previously known — among them the multitude of small fry whose existence is indicated by the reionization data. Collection of data from the first four 'frontier fields' is scheduled to be completed over the next two years, says Mark Dickinson of the National Optical Astronomy Observatory in Tucson, Arizona.

In Chile, ALMA will join the hunt for distant galaxies from this summer (see *Nature* 495, 156–159; 2013). In contrast to Hubble, which records starlight, ALMA's microwave measurements will reveal the gas and dust that gives rise to stars in these remote bodies. Paradoxically, says James Dunlop of the University of Edinburgh, UK, a member of the 2012 HUDF team, this will allow ALMA to make the most accurate measurement yet of starbirth at such distances. Newborn stars radiate most of their light at ultraviolet wavelengths, he explains, but much of that light is absorbed by gas and dust and reradiated at infrared wavelengths, which are then redshifted by cosmic expansion into ALMA's millimetre range.

In addition, ALMA's high spatial resolution will enable the array to break radio emissions into their component wavelengths and therefore record the actual redshifts — bona fide measurements of distance — of many of the remote galaxies that Hubble has studied, says Chris Carilli at the National Radio Astronomy Observatory in Socorro, New Mexico.

Astronomers can then translate those distance measurements into ages, which will give them a much better handle on where these objects fit in cosmic history.

"Hubble has been amazing at finding candidate galaxies from redshift 7 to 10, but none of these has been confirmed with spectra and the potential for [spurious candidates] is severe," says Carilli.

Carilli and his collaborators reported<sup>7</sup> in February that ALMA can measure redshift-7 galaxies (objects 3,955 megaparsecs or 12.9 billion light years from Earth) using just 20 of its eventual complement of 66 antennas. A report from another team in *Nature*<sup>8</sup> provides further evidence. ALMA will "quickly make the jump to redshift 8" by the end of the year, says Carilli, and if the array gets a new set of receivers — a possibility still several years in the future — it could study and measure distances to galaxies out to redshift 11. These objects would be seen as they appeared just 425 million years after the birth of the Universe. ALMA could become the "redshift machine of choice" for the first galaxies, he says.

Nevertheless, most astronomers are eagerly awaiting the 6.5-metre JWST, whose *raison d'être* is to image the faint, primitive bodies that Hubble can only glimpse — but that were the earliest ancestors of modern galaxies such as the Milky Way (see *Nature* 467, 1028–1030; 2010). Hubble's observations provided "the first hints of the first galaxies", says Ellis, but "we really need the JWST to push back into that even earlier period from 200 million years to 500 million years after the Big Bang".

Back in 2009, even as NASA astronomer and astronaut John Grunsfeld glimpsed the first images of distant galaxies from the infrared camera he had helped to install on Hubble, the JWST had come to his mind. "I couldn't help but feel awed by the power of Hubble," Grunsfeld recalls, but "the views of the HUDF also gave me great satisfaction that the JWST would have plenty to see." ■

**Ron Cowen** is a freelance writer in Silver Spring, Maryland.

1. Bouwens, R. J. *et al.* <http://arxiv.org/abs/0909.1803> (2009).
2. Oesch, P. A. *et al.* <http://arxiv.org/abs/0909.1806> (2009).
3. McLure, R. J. *et al.* *Mon. Not. R. Astron. Soc.* <http://dx.doi.org/10.1093/mnras/stt627> (2013).
4. Robertson, B. E. *et al.* *Astrophys. J.* **768**, 71 (2013).
5. Schenker, M. A. *et al.* *Astrophys. J.* **768**, 196 (2013).
6. Bouwens, R. J. *et al.* *Astrophys. J. Lett.* **752**, L5 (2012).
7. Wang, R. *et al.* <http://arxiv.org/abs/1302.4154> (2013).
8. Vieira, J. D. *et al.* *Nature* **495**, 344 (2013).

# COMMENT

**ENVIRONMENT** Street lights that use less energy are better for people and animals **p.560**

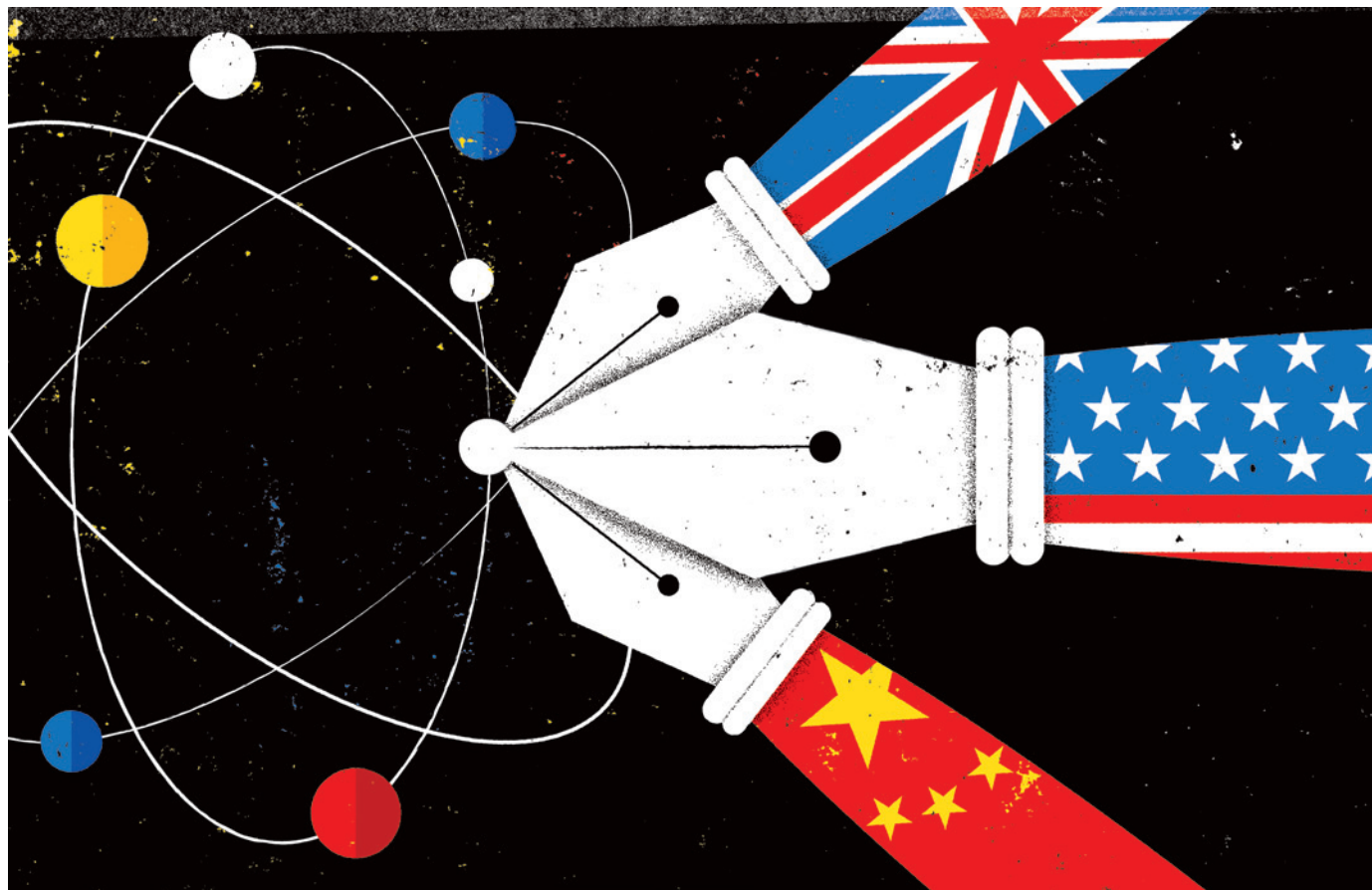
**BIOGRAPHY** A life of Nikola Tesla — engineer, inventor and showman **p.562**

**SPACE** Astronomer at Royal Observatory talks celestial images **p.564**



**GENETICS** Time for biologists to venture beyond the central dogma **p.565**

ILLUSTRATION BY DALE MURRAY



## The fourth age of research

**Jonathan Adams** analyses papers from the past three decades and finds that the best science comes from international collaboration.

Research has progressed through three ages: the individual, the institutional and the national. Nations competed to be at the cutting edge because this contributed to the wider economy through knowledge, new processes and products.

Today, we are entering a fourth age of research, driven by international collaborations between elite research groups. This will challenge the ability of nations to conserve their scientific wealth either as intellectual property or as research talent<sup>1–3</sup>. Tensions are

growing: between the knowledge a country needs to remain competitive and the assets it can exclusively secure, and between the collaborative and domestic parts of the research base. Institutions that do not form international collaborations risk progressive disenfranchisement, and countries that do not nurture their talent will lose out entirely.

To explore the scale of this challenge, I analysed data on research articles and reviews from Thomson Reuters Web of Science between 1981 and 2012. I included

papers in a country's tally if one or more author addresses included that country, but no weighting was applied for the number of authors or addresses on each paper. For the 25 million papers I included in my analysis, I calculated the balance of international and domestic research collaboration for established economies (the United States, the United Kingdom, Germany, France, the Netherlands, Switzerland) and emerging ones (China, India, South Korea, Brazil, Poland). I looked particularly at all ►

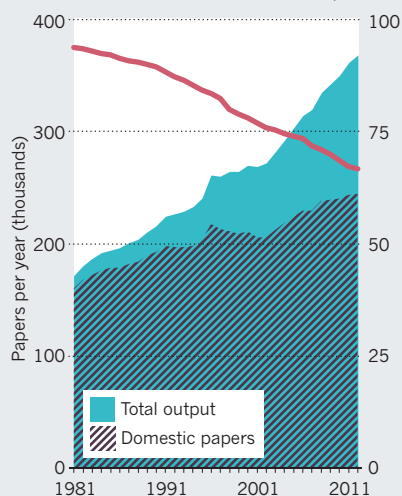


## STRENGTH IN NUMBERS

Growth in international collaboration eclipses domestic output in established economies, but not in emerging ones.

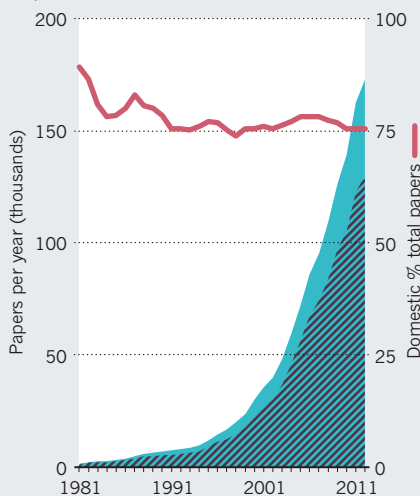
### UNITED STATES

The country is less internationally collaborative than those in Western Europe.



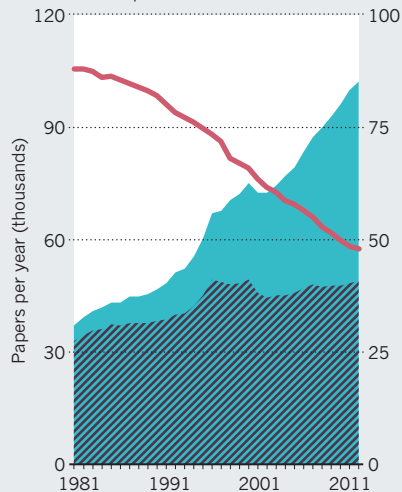
### CHINA

More than three-quarters of research output remains domestic.



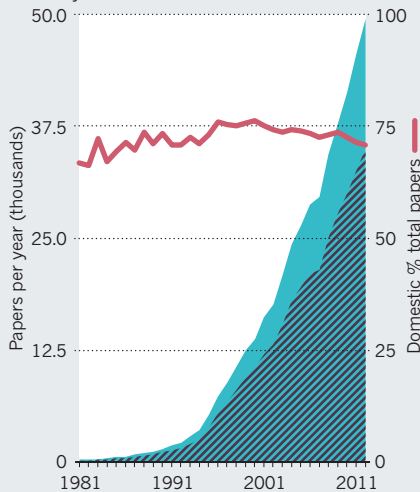
### UNITED KINGDOM

International collaboration has almost doubled in the past decade.



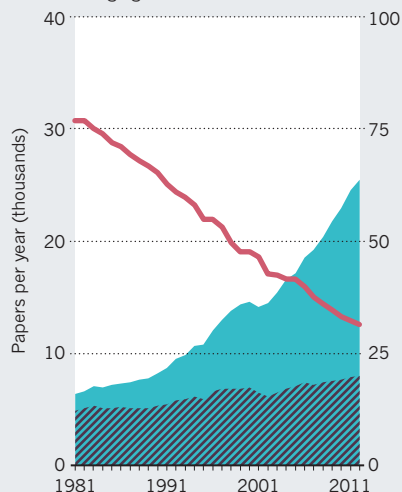
### SOUTH KOREA

Even more rapid growth than China, driven by domestic research.



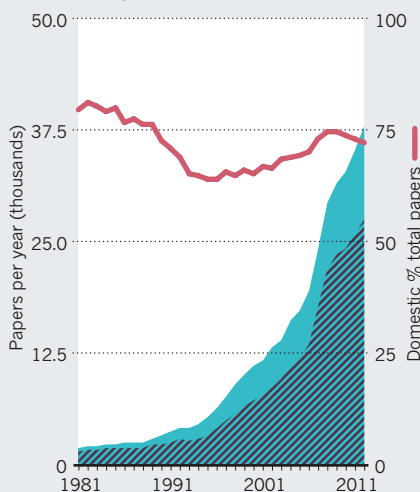
### SWITZERLAND

Multinational programmes have led to above-average growth and citations.



### BRAZIL

Collaboration varies but domestic research is as predominant as in Asia.



► papers from the United Kingdom and the United States that listed multiple authors. Then I probed the relative citation impact of domestic and international research in the two countries.

I found that, over more than three decades, domestic output — papers that list only authors from the home country — has flatlined in the United States and in Western European countries. The rise in total annual output for each country is due to international collaboration. As a result, the percentage of papers that are entirely 'home grown' is falling. In emerging economies, by contrast, domestic output is rapidly expanding (see 'Strength in numbers').

These shifts stand to change the dynamics of how nations invest in science, who owns its outputs and who best exploits them.

## BORDER CROSSING

In Western Europe in 1981, domestic share of output was about five papers in every six published by each country, but fell to half of total output in 2011. The United States started as less collaborative — just 6% of US papers in 1981 had a foreign co-author, but it too has had a marked fall in domestic share of output: it is the most frequent partner for many other nations. Nonetheless, it remains less internationally collaborative than Western Europe.

In emerging economies in 1981, the domestic share of output was similar to that of the established economies. Today, about 75% of the research output of China, Brazil, India and South Korea remains entirely domestic. The total volume of papers from these four countries has increased 20-fold — from fewer than 15,000 papers annually in 1981 to more than 300,000 papers now. Despite its European location, Poland's balance of domestic and collaborative papers is closer to this group (see Supplementary Information at [go.nature.com/nszcek](http://go.nature.com/nszcek)). Quantity is now being matched by significant improvements in quality: the citation impact of more than 10% of China's domestic research is already above twice the world average.

For established economies, total national research output has more than doubled over the past 30 years. Meanwhile, domestic output has increased by only about 50% for all except the Netherlands (which more than doubled). After the mid-1990s, the domestic research output of the United Kingdom (47,500 papers per year), Germany (45,000 papers) and France (30,000 papers) levelled off while international collaboration in these countries increased more than ten-fold.

What fuelled this growth in collaboration? Much of it has come from bilateral partnerships rather than multinational programmes (see Supplementary Information). For instance, less than 1% of UK papers from 2003 to 2012 list co-authors from all three of France, Germany and the United States, and

SOURCE: THOMSON REUTERS WEB OF SCIENCE; ANALYSIS: EVIDENCE

less than 0.1% of US output over the same period has co-authors from the United Kingdom, Germany and China together.

Switzerland is an interesting outlier. For a small country, its research output has seen above-average growth and a high average citation impact. Today, more than two-thirds of its publications have a foreign co-author. A large part of this comes from the country's hosting of multinational programmes, such as CERN (Europe's particle-physics lab) and the World Health Organization, both in Geneva.

Citation impact is typically greater when research groups collaborate, and the benefit strengthens when co-authorship is international<sup>4-6</sup>. I calculated the citation impact of each UK and US paper published in 2001 and then for 2011 by normalizing citation count to take into account the time elapsed since year of publication and the subject area of the journal. For each of these years, I then calculated an average normalized citation impact (where 1.0 is the world average) for papers that were purely domestic and for papers with a foreign co-author.

For both countries, papers that are internationally collaborative are cited relatively more often than purely domestic papers. And this 'impact premium' rose in both countries by around 20% between 2001 and 2011 (see 'Impact premium').

### A CASE IN POINT

A closer look at the United Kingdom reveals the drivers of some of these changes. The UK government's Department for Business, Innovation & Skills (BIS) reports both rising volume and rising average citation impact for the UK research base relative to the world average. My analysis shows that this impact (up from 1.21 in 2001 to 1.47 by 2011) is mainly attributable to the near-doubling of its international collaboration (from 26,608 to 50,423 papers with an average impact up from 1.48 to 1.72). By comparison, the UK domestic research base, which BIS has never analysed separately, has changed much less in quantity and quality. In 2001, the country published 46,671 domestic papers with an average impact of 1.05; in 2011 it published 47,323 domestic papers with an average impact of 1.21. And in 2010, the number of UK papers with no overseas co-author actually fell below the count of internationally collaborative papers.

The benefits of collaboration are systematic. I found that 130 UK universities and colleges published more than 100 research articles and reviews in journals indexed on Thomson Reuters Web of Science in 2002–11. Around half of these universities have an international co-author on less than 40% of their papers. Average impact in this group rarely gets much above 1.1. By contrast, research staff in institutions with greater than 50% international

### IMPACT PREMIUM

In the United States (US) and the United Kingdom (UK), papers with at least one author from another country are cited more often than purely domestic work.



co-authorship on their papers had an average citation impact above 1.6 (see Supplementary Information). This category includes the five or six UK universities that are generally recognized as standing ahead of others, including the universities of Oxford and Cambridge.

This same handful of elite institutions receives about one-third of all available core research income from the UK higher-

education funding councils, more than one-third of the total available research grant and contract income and one-third of all research-council studentships. By contrast, the 65 or so UK universities that have an international co-author on less than 40% of their papers collectively account for less than 10% of funding-council research money, less than 5% of research grants and contracts and less than 1% of research-council studentships (my analysis; using data from the UK Higher Education Statistics Agency).

**"Elite national universities are also leading international collaborators."**

### CONTEMPORARY GIANTS

First-rate researchers are, and always were, thinly distributed. In the past they could collaborate only intermittently. Contemporary giants call more readily across the desolate intervals of space (to paraphrase Friedrich Nietzsche) and rapidly receive answers, because information technology has sharply reduced the costs of collaboration<sup>7</sup>.

Excellence seeks excellence, so elite national universities are also leading international collaborators. Exceptional research groups share ideas, resources and outcomes. For example, the most frequent international partners of the University of Cambridge, UK, are the Max Planck institutes in Germany, the Massachusetts Institute of Technology and Harvard University, both in Cambridge; the

California Institute of Technology in Pasadena; the University of California, Berkeley; and the universities of Toronto, Heidelberg and Tokyo. Harvard's frequent international partners are Imperial College London, University College London, the Max Planck institutes, the Karolinska Institute in Stockholm and the universities of Cambridge, Toronto and Geneva. Internationally co-authored papers are more highly cited because the authors are more likely to be doing excellent research. Multinational programmes are still marginal as drivers of performance.

So, in this fourth age of research there is a growing divide between international and domestic research. This will influence each nation's ability to draw on the global knowledge base, and could in turn compromise national scientific wealth. For established economies, the risks are a gradual financial and intellectual separation between institutions that are largely international and those that are largely national. Such separation could lead to the erosion of adequate regional competency for future research training and collaboration and for knowledge flow to the national industrial base.

Government policy should focus on three issues to address these risks. First, in this age of big data that are internationally shared, the question will be who has the skills to exploit knowledge assets fastest, not who owns them. Shared knowledge and discovery sideline the idea of securing intellectual property by commercial confidentiality and patenting. The growing scarcity of truly able researchers could lead to a 'global war for talent'<sup>8</sup>. So governments must provide and sustain conditions that attract and support the best scientists, otherwise talent flows elsewhere and there is no capacity to generate replacements.

Second, incentives must be put in place to enable universities to participate in international networks. For example, tangible projects involving participation could be ►



► given explicit recognition and credit in systems for assessing research quality (such as the UK Research Excellence Framework). Insight into the evolving research of the emerging economies will be limited without active engagement and collaboration. Relying simply on what is published in journal articles has long been recognized as a poor mode of knowledge transfer<sup>9</sup>.

Third, collaboration must include European and US researchers in Asian and South American laboratories. The older economies can no longer rely on the best foreign researchers to come and visit. Maintaining a talent pool is necessary, but it is not sufficient: if a nation really wants to remain at the highest level scientifically, it needs to get its talent out and about — and then bring it back again. One way of doing this is to create worthwhile travel and exchange programmes that include an assured post for the returning researcher. The United Kingdom has a shocking record on mobility in European research programmes and the United States has the lowest level of international collaboration among the G7 countries. By contrast, the Netherlands achieves excellence by enabling its researchers to be much more mobile, despite its lesser resources<sup>10</sup>.

Impact and innovation will flow from a coalition of the willing, not the straitjacket of international policy and coordination. Multinational programmes are not the answer, nor are academic memoranda of understanding. To maintain the dividend that governments garner from research excellence, they must ensure that universities and their researchers have the resources, facilities and incentives to create and sustain flourishing partnerships. ■

**Jonathan Adams** was formerly director of research evaluation at Thomson Reuters and is now a research-management consultant in Leeds, UK. e-mail: jonathanzadams@gmail.com

1. May, R. M. *Science* **275**, 793–796 (1997).
2. Adams, J. *Nature* **396**, 615–618 (1998).
3. Adams, J. *Nature* **490**, 335–336 (2012).
4. Van Raan, A. F. J. *Scientometrics* **42**, 423–428 (1998).
5. Adams, J., Gurney, K. A. & Marshall, S. *Scientometrics* **72**, 325–344 (2007).
6. Schmoch, U. & Schubert, T. *Scientometrics* **74**, 361–377 (2008).
7. Adams, J. D., Black, G. C., Clemmons, J. R. & Stephan, P. E. *Res. Pol.* **34**, 259–285 (2005).
8. Dobbs, R., Lund, S. & Madgavkar, A. 'Talent tensions ahead: a CEO briefing.' *McKinsey Quarterly* (November 2012); available at <http://go.nature.com/hz7ars>.
9. Gibbons, M. et al. *The New Production of Knowledge* (Sage, 1994).
10. Gurney, K. & Adams, J. *Tracking UK and International Researchers by an Analysis of Publication Data* (Higher Education Policy Institute, 2005).



Wasted light from street lights, such as those in New York City, contributes to skyglow.

# A green light for efficiency

Efforts to improve street lights are providing a rare opportunity to cut both financial and environmental costs, argues **Kevin Gaston**.

Over the past few years, local governments in towns and cities across the world have been dimming street lights, turning them off for parts of the night, investing in more energy-efficient technologies and gradually modernizing old lighting systems. Hard times and tight budgets have helped to drive this flurry of activity.

The combination of financial pressures, a broad array of new lighting technologies and a richer understanding of the importance of natural-light cycles to the health of many organisms presents an unusual opportunity. To take advantage, environmental scientists must work quickly to gather and build on existing knowledge about the effects of artificial light on plants and animals, and must collaborate with those responsible for

the development and implementation of public lighting schemes.

As well as disrupting the development, physiology and behaviour of many organisms, grid-based electric lighting is estimated to produce 1.5 billion tonnes of carbon dioxide each year globally. Moreover, artificial light disturbs people's natural circadian rhythms. Various studies have linked this disturbance to conditions such as breast cancer<sup>1</sup> and obesity. A notorious impediment to astronomical observations, artificial light also robs us of moonlit landscapes and properly dark or star-filled skies.

Most existing street-lighting systems are inefficient and expensive. Much of the light they produce is wasted, contributing to skyglow often tens and possibly hundreds of





kilometres away. Historically, some lighting systems were installed not to aid people, but because they were thought to add to the beauty of a city, or because they could be run at little extra cost by using excess electricity produced during periods of low demand. Likewise, the illumination of many roads has not been adjusted to account for the fact that headlights on modern vehicles are now some eight times brighter than those of cars and trucks built 60 years ago.

### DIM THE LIGHTS

A global assessment<sup>2</sup> of artificial lighting, carried out in 2006, estimated that street lighting consumes about 114 terawatt hours of electricity globally each year. For comparison, this is almost 400 times the electricity that New York City consumes for street lighting annually. National and local governments in the United States, Europe and Asia have quickly recognized the potential for savings — not just by dimming and turning off street lights, but also by installing energy-efficient light-emitting diodes (LEDs) instead of conventional bulbs. As the cost of new technologies continues to fall, such efforts will increase.

There are various ways to cut the energy used by street lights. These generally involve reducing the time or extent to which lights are used, and their intensity<sup>3,4</sup>. Savings depend on the size and type of

existing lighting systems, their purpose and whether systems are being newly installed or upgraded. If new roads, towns or cities are being built, for instance, investing in low-energy and centrally manageable lighting makes sense. The feasibility of retrofitting existing developments depends in part on how long it takes for the long-term savings in energy costs to cover the expense of replacement. For LED technology, this is currently about 10–20 years.

Those responsible for modifying lighting systems should consider other factors too. A nuanced approach would ensure that lighting systems do the least environmental harm, as well as providing the necessary benefits at minimal expense.

Artificial light can influence myriad biological processes, including sleep, metabolism, germination and flowering. It can alter the distribution and abundance of organisms and disrupt predator–prey relationships<sup>5,6</sup>. In fact, such effects are probably pervasive, given the importance of natural-light cycles on circadian clocks and organisms' physiological responses to day length.

In general, limiting the number of street lights, dimming them and turning them on for shorter periods reduces these negative biological effects<sup>6</sup>. But studies are needed to identify the best compromises. People are most in need of artificial light in the hours after dusk and before dawn — just when other species are most likely to perceive the day to be extended (if diurnal) or curtailed (if nocturnal). Adjusting these periods to better suit the needs of local ecology, within the constraints of human requirements, could reduce the negative impact of night-time lighting all round.

White-light LED systems and other white-light technologies are proving popular because they improve colour rendering for humans — objects look more as they would when viewed in natural light. But increasing the range of wavelengths of a light source also increases the overlap between the emitted light and the spectral sensitivities of a broad array of other organisms. Some white lights emit a lot of ultraviolet, making them more attractive to insects such as moths. These technologies should be limited to areas where they will be most useful and ecologically least destructive, such as in urban rather than rural settings. Parts of the light spectrum that are not useful to people should be avoided.

### EFFICIENCY AND ECOLOGY

In Europe, major projects to improve the evidence base for ecologists' recommendations for lighting schemes are under way. These include the ECOLIGHT project, an investigation of the ecological effects of light pollution, which I lead at the University of Exeter, UK. The experimental components of this work need to be expanded to investigate the

environmental effects of night-time lighting in a sufficiently broad array of ecosystems and taxa.

To ensure that everyone wins from the growing worldwide movement to reduce night-time lighting, a three-way dialogue

**“Artificial light can influence sleep, metabolism, germination and flowering.”**

is vital between those investing in new lighting technologies, those producing them and environmental scientists. My group and I have found that discussions with local authorities enable us to better envision and articulate how social

challenges can be weighed against environmental ones. Equally, our findings offer such authorities more leverage to bring about changes.

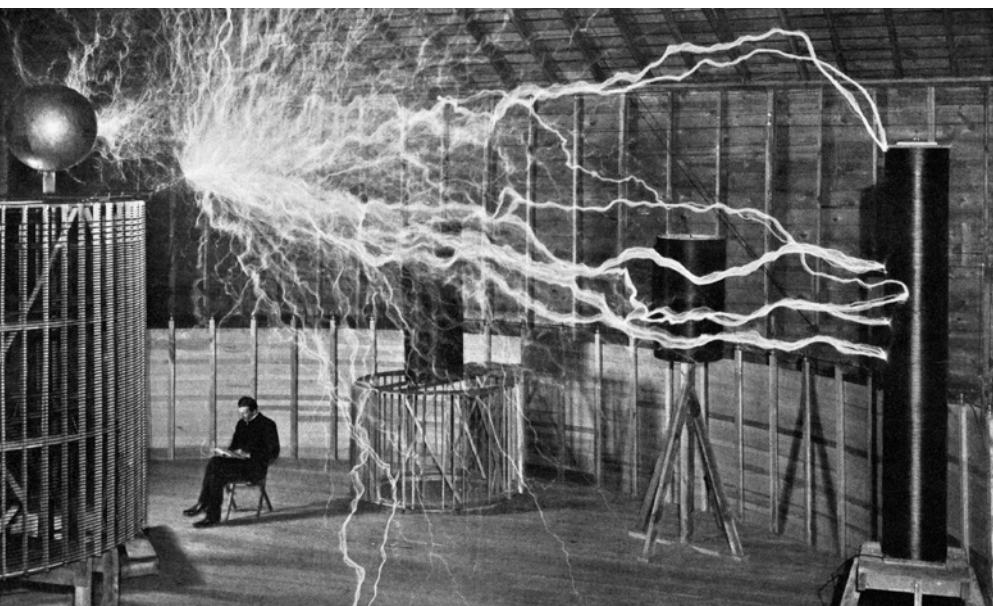
Through these discussions I have come to believe that the biggest barrier that local and national governments face in their efforts to reduce night-time lighting is the public perception of its importance in reducing vehicle accidents and crime. Unsurprisingly, almost all of the evidence connecting the importance of lighting to safety and security comes from experimental studies conducted in extremely low-risk areas, or correlational studies that are notoriously difficult to interpret<sup>7,8</sup>.

Determining the effects of lighting is also complicated by the fact that people alter their behaviour in response to lighting changes<sup>9</sup>. Brighter street lights, for example, can make people feel more confident and therefore induce them to drive faster. Thus, in addition to studies on the ecological impact of different lighting regimes, research is needed to investigate their societal impact.

This opportunity to install street lighting that uses less energy and is better for human, animal and plant health must be grasped now. It may not arise again for many decades. ■

**Kevin J. Gaston** is director of the *Environment and Sustainability Institute, University of Exeter, Penryn, UK.*  
e-mail: k.j.gaston@exeter.ac.uk

1. Stevens, R. G. *Int. J. Epidemiol.* **38**, 963–970 (2009).
2. International Energy Agency. *Light's Labour's Lost: Policies for Energy-Efficient Lighting* (International Energy Agency, 2006).
3. Falchi, F., Cinzano, P., Elvidge, C. D., Keith, D. M. & Haim, A. *J. Environ. Mgmt* **92**, 2714–2722 (2011).
4. Gaston, K. J., Davies, T. W., Bennie, J. & Hopkins, J. *J. Appl. Ecol.* **49**, 1256–1266 (2012).
5. Rich, C. & Longcore, T. (eds) *Ecological Consequences of Artificial Night Lighting* (Island Press, 2006).
6. Gaston, K. J., Bennie, J., Davies, T. W. & Hopkins, J. *Biol. Rev.* (in the press).
7. Marchant, P. R. *Rad. Stat.* **102**, 32–42 (2010).
8. Beyer, F. R. & Ker, K. *Cochrane Database of Systematic Reviews* <http://doi.org/cxj8mk> (2010).
9. Assum, T., Bjørnskau, T., Fosser, S. & Sagberg, F. *Accid. Anal. Prev.* **31**, 545–553 (1999).



Artificially generated lightning crackles around Nikola Tesla's Colorado laboratory.

## PHYSICS

# The mind electric

Patrick McCray assesses a biography of Nikola Tesla, the Serbian wizard of the alternating current.

When entrepreneur Elon Musk named his all-electric car company Tesla Motors, he was paying homage to a remarkable man. Serbian inventor and electrical engineer Nikola Tesla (1856–1943) created a veritable zoo of electrical inventions, from motors that used alternating current (AC) to radio-controlled boats, and a proposed system for the wireless transmission of electricity from one continent to another.

Bernard Carlson's superb biography follows Tesla from his years in what is now Croatia, Austria and Hungary, where he studied physics, engineering and mathematics, to his 1884 arrival in New York City and then on to spectacular successes and failures in electrical innovation. Carlson brings to life Tesla's extravagant self-promotion, as well as his eccentricity and innate talents, revealing him as a celebrity-inventor of the 'second industrial revolution' to rival Thomas Alva Edison.

Tesla worked briefly for Edison in the United States, but quit in disgust when Edison declined to use his arc-lighting system. Financially strapped, Tesla plunged into the technological ecosystem of New York, with its growing demand for electricity for the power and communications industries. As Carlson relates, these technological frontiers were largely unsettled at the time, with inventors,

entrepreneurs and financiers all fighting to secure a competitive advantage. A key question was whether Edison's direct-current approach or the AC option favoured by Tesla would come to dominate power transmission.

Tesla's signal accomplishment was the realization of his AC motor in the late 1880s. This was based on a series of inventions and patents for using two sources of alternating current out of phase with one another. Tesla saw that these created a rotating magnetic field that could make a motor. His polyphase power systems were backed by funds from US entrepreneur George Westinghouse and ultimately allowed utilities to transmit electrical power over longer distances than before.

The first biography on Tesla appeared in 1894, when he was riding high after building an "oscillating transmitter" (a resonant transformer also called the Tesla coil) and seeing his polyphase AC motors deployed for power generation at Niagara Falls. Public lectures in the United States and Europe followed. Tesla had a consummate ability to impress audiences and potential backers with stunning electrical displays. An 1891 lecture, for instance, featured two large zinc sheets suspended from the

auditorium ceiling and connected to a power source. Under dim lighting, Tesla stepped between the slabs holding a gas-filled tube in each hand. The electrostatic field made the tubes glow as Tesla told the enthralled crowd how electric lights might be moved around without being tethered to wires.

Tesla further studied the possibility of wireless power transmission in 1899, while on sabbatical in Colorado Springs. (A well-known double-exposure photograph, **pictured**, in which he seems to sit nonchalantly amid fierce electrical discharges, was taken here.) The region's summer storms led him to surmise that lightning bolts initiated electromagnetic waves in Earth's crust, creating stationary waves. He believed that this process would allow power transmission "in unlimited amounts, to any terrestrial distance and without loss".

Many writers have cast Tesla as a "Nietzschean superman", as Carlson puts it. But Carlson looks critically at Tesla's wilder claims — such as his proposed particle-beam weapon, which never came to fruition despite stirring up interest among Soviet, British and US officials on the eve of the Second World War. Carlson is also frank about Tesla's misunderstanding of scientific discoveries. In 1887, for example, German physicist Heinrich Hertz detected the electromagnetic waves predicted by Scottish physicist James Clerk Maxwell. Tesla decided that the glow in evacuated glass discharge tubes was due to "electrostatic thrusts". The error did not prevent him from trying to convert Hertz's basic discovery into devices in the lab. Technological innovation, not scientific discovery, remained this visionary engineer's primary goal.

Carlson contextualizes Tesla's approach with excursions into the nature and psychology of invention, exploring theories such as economist Joseph Schumpeter's model of "creative destruction" and business professor Clayton Christensen's more recent idea of "disruptive innovation". Both these theories are based on the idea that entrepreneurs and inventors who produce radical technologies can cause widespread social and economic disruption. Tesla's inventive style, says Carlson, existed in "tension between ideal and illusion": he first shaped inventions in his mind rather than taking Edison's empirical approach. Tesla believed that his polyphase system rested on a beautiful principle, which he expected businessmen and customers to adapt to. However, putting design ideals ahead of practical considerations — not unlike Apple's late chief executive Steve Jobs, Carlson notes — sometimes meant that Tesla missed out on commercial possibilities.

In about 1900, Tesla began to conceptualize a grand system that would enable the wireless transmission of power and communications "from Pike's Peak" in Colorado's Rocky Mountains "to Paris". Buoyed by support from

**Tesla: Inventor of the Electrical Age**  
W. BERNARD CARLSON  
Princeton University Press: 2013. 520 pp.  
£19.95, \$29.95



Wall Street wizard John Pierpont Morgan, he oversaw the building of Wardencllyffe, an imposing laboratory-cum-transmitting tower on Long Island, New York. However, faulty research results undermined the plan and, unnerved by Tesla's self-promotion and suspicious of the growing speculative bubble in wireless communications, Morgan withdrew. The giant transmission facility was never completed and Tesla found himself scrambling for cash. (Despite this, Tesla's activities probably spurred physicist and inventor Guglielmo Marconi to speed up his own work on wireless communications.)

The failures that dogged Tesla's transmission scheme bit deeply into his perceptions of how the electrical world worked. Earth didn't behave as if it were filled with an incompressible fluid, as Tesla believed. When the Wardencllyffe experiment failed, Carlson recounts, Tesla faced a "serious dilemma ... Either he was wrong or nature was wrong." Ideas clashed with reality and Tesla, angry and depressed, had a nervous breakdown in 1905.

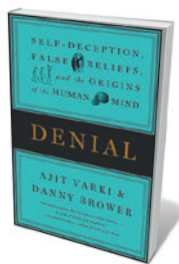
Tesla's last three decades in New York City were spent in relative obscurity. He never gave up the dream of wireless power transmission and continued to invent while living precariously off a modest stream of royalties. He gave annual press conferences in which he speculated about the future of technology. His life, one observer noted, took on a more "speculative, philosophical, and somewhat promotional character".

In the early 1970s, many years after his death, Tesla's enigmatic behaviour (such as his passion for feeding pigeons) and lingering reputation for grandstanding electrifying illusions helped to cement his appeal among free-energy claimants — who believe there is electricity in our environment waiting to be picked up with the right technology — and counterculturalists seeking mystery in the rational and material world. Public interest has spiked again in the past few years. In 2012, an online campaign raised close to a million dollars in a week for a Tesla museum. A recent YouTube video has Tesla duelling with Edison in a rap battle, and Christopher Nolan's 2006 film *The Prestige* features David Bowie as Tesla. Tesla's bold predictions and outsider glamour are still working their magic.

Tesla shows that a deep creative drive, guided by a formidable intuition, can serve inventors well up to a point. Innovation may be a subjective process, but Tesla's career demonstrates that it can also be undermined by believing that illusion has substance. ■

**W. Patrick McCray** is a professor in the Department of History at the University of California, Santa Barbara. His latest book is *The Visioneers: How a Group of Elite Scientists Pursued Space Colonies, Nanotechnologies, and a Limitless Future*. e-mail: pmccray@history.ucsb.edu

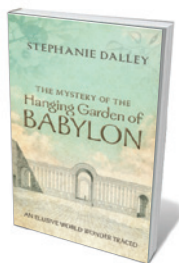
## Books in brief



### Denial: Self-Deception, False Beliefs, and the Origins of the Human Mind

Ajit Varki and Danny Brower TWELVE 384 pp. \$27 (2013)

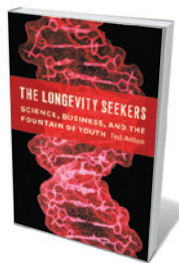
Do you skydive? Deep-fry? Chain-smoke? Denial of mortality is a strange trait that is also key to human nature, argues medic Ajit Varki. His argument stems from the ideas of late geneticist Danny Brower, who asked why species such as chimpanzees have not evolved to be aware of both self and the minds of others. Varki speculates that such intersubjectivity could only arise in tandem with 'death blindness', as fear would otherwise hamstring a species' fitness. A thoughtful foray into "mind over reality".



### The Mystery of the Hanging Garden of Babylon: An Elusive World Wonder Traced

Stephanie Dalley OXFORD UNIVERSITY PRESS 304 pp. £25 (2013)

The puzzling dearth of research on Babylon's Hanging Garden, one of the seven wonders of the ancient world, prompted Assyriologist Stephanie Dalley to methodically sift the evidence. Her perusal of cuneiform tablets, rock reliefs and Latin texts yielded research gold, overturning long-held ideas about the creator and location of this vertiginous marvel. From its fantastical landscaping to its advanced irrigation system, the garden emerges as a wonder indeed — of engineering, aesthetics and metaphoric richness.



### The Longevity Seekers: Science, Business, and the Fountain of Youth

Ted Anton UNIVERSITY OF CHICAGO PRESS 240 pp. \$26 (2013)

A "silver tsunami" is upon us, writes Ted Anton: by 2050, one-third of people in the developed world will be over 60. The time has come to tease out the "molecular tipping points" involved in maintaining geriatric health, Anton avers. Kicking off with molecular biologist Cynthia Kenyon — who in 1993 pinpointed a single-gene mutation that doubles the lifespan of the nematode *Caenorhabditis elegans* — Anton reveals a young field already rife with larger-than-life personalities and lab drama aplenty.



### Kith: The Riddle of the Childscape

Jay Griffiths HAMISH HAMILTON 432 pp. £20 (2013)

Sojourning on several continents to research her bestselling *Wild*, Jay Griffiths noted big differences between children from indigenous and Western cultures. She now grapples with that riddle, arguing that "human nature is nested in nature which co-creates the child". Her probings of the meeting point of developing psyche and environment interweave history, anthropology and memoir. But does an urban existence enfeeble the young? What is abundantly clear, yet sidelined in this often brilliant, poetically nuanced work, is the ferocious adaptability of our species and our children.



### Billion-Dollar Fish: The Untold Story of Alaska Pollock

Kevin M. Bailey UNIVERSITY OF CHICAGO PRESS 288 pp. \$25 (2013)

The last time you ate something labelled just 'fish', it might well have been pollock. The flesh of this Alaskan species turns up in fish fingers, sushi and seafood salad. Kevin Bailey, a former senior scientist at the Alaska Fisheries Science Center in Seattle, presents the first natural history of this ubiquitous fish and an analysis of its population. Although the market for pollock — worth more than a billion dollars a year in the United States alone — seems buoyant compared with some others, Bailey unveils a familiar tale of steep decline. **Barbara Kiser**





An image of the Butterfly Nebula captured by the Hubble Space Telescope.

## Q&A: Marek Kukula

# Star-map historian

Marek Kukula is public astronomer at London's Royal Observatory in Greenwich and the curator of *Visions of the Universe*, an exhibition charting the trajectory of celestial imaging, with a focus on astrophotography. On the eve of its opening, Kukula talks about eighteenth-century star maps and the co-evolution of the telescope and camera.



### How did you get into astronomy?

I decided I wanted to be an astronomer during a family trip to Jodrell Bank in Cheshire, UK, when I was in my teens. I remember looking up at the giant radio telescope and learning about how scientists were using it to try to answer some of the biggest questions in the Universe. It seemed like a huge intellectual adventure.

### What links the Royal Observatory to the exhibition's venue, the National Maritime Museum?

The observatory was built to make very accurate maps of the stars for use in navigation, which is why astronomy had government funding in the seventeenth and eighteenth centuries. It was really only in the nineteenth century that people began to use observatories for modern astronomy and astrophysics.

### What is the idea behind *Visions of the Universe*?

The focus is astrophotography and its evolution, told through roughly 100 iconic images of space, primarily from the past 150 years [though there are also some sketches from the pre-photography era]. The role of the telescope in astronomy is familiar, but a parallel story has been less well told: the impact of the invention and development of the camera on this field. We wanted to reveal how the telescope and the camera have changed the way we see the Universe, a story astronomers can be very proud of. Although the earliest cameras were not really suitable because they needed quite a lot of light, astronomers were among the first people to recognize their potential, and helped to develop the technology to a point at which it was actually useful.

### How did you choose the images?

It has been a really agonizing process trying to narrow them down. We tried to pick ones that tell a good story, and included some because they just look so beautiful. Some are hugely important in scientific terms, if not in aesthetic ones. The first image beamed back

**Visions of the Universe**  
National Maritime Museum, London.  
7 June to 15 September 2013.

in 1959 by one of the Soviet lunar probes as it went behind the Moon is a really grainy television picture. When we dragged it up on the Internet, we burst out laughing because it is such a rubbish image. But it is hugely significant, both scientifically and culturally. It was the first time anyone had seen the other side of the Moon.

Other images combine artistry with science. There is a really famous one from the Cassini space probe of Saturn eclipsing the Sun. There's a technology story here: this is of course a view that's impossible from Earth. Because we can send a camera on a spacecraft out to Saturn, we're able to see this really stunning view. Between the inner and outer rings is a tiny speck, which is Earth. We're in the picture as well.

### Which are your favourite images?

One is of the Andromeda Galaxy taken by Edwin Hubble on 19 October 1923. At the time, there was a great debate about whether the Milky Way was the entire Universe. Hubble picked out one star that is variable in brightness, so he wrote 'VAR!' on the photographic negative. The key thing about variable stars is that if you can measure their variability, you can calculate their distance. When he worked this out, he discovered that the Andromeda Nebula is 2.5 million light years away, well beyond the Milky Way. This image revealed the Universe as hundreds of times bigger than we had assumed.

### What's the thinking behind the pictures from the surface of Mars?

I find those images incredibly moving. On one level they look ordinary and boring: there's a desert; there are some rocks. But they're on another planet — that's incredible! For thousands of years, people saw Mars as a red dot in the sky. About 400 years ago, they started to see it as a tiny disc and began to see landscape features. We are the first generation in history to see it as a landscape, with pebbles and sand dunes and hills, and to drive our robot avatar through it.

### What do you enjoy most about being public astronomer at the Royal Observatory?

I get to work alongside historians of science. At university you tend to see science as a linear progression, but the observatory's curators remind me that scientists in earlier centuries had no idea of the correct route to pursue. That gives you a much better understanding of why we do things in sometimes quite perverse ways in modern science: this is the legacy of how we've arrived at our modern understanding.

INTERVIEW BY DANIEL CRESSEY

NICK HIGGINS

# Correspondence

## Account for the 'dark matter' of biology

Philip Ball correctly opines of DNA that "we should lift some of the awesome responsibility for life's complexity from its shoulders" (*Nature* **496**, 419–420; 2013). In addition to the '-omes' he lines up for this responsibility, other components that are vital to all life forms are the metabolome (small-molecule intermediates), the lipidome (including cellular membranes) and the glycome (massive arrays of sugar chains on cell surfaces and extracellular compartments).

We now know that these other '-omes' mediate much of the diversity and complexity found in natural biological systems. Ball makes an apt analogy to the 1998 discovery that the expansion of the Universe is accelerating, which forced cosmologists to think beyond the standard model and realize that dark matter and dark energy actually dominate many key processes.

Likewise, it is time for more biologists to venture beyond the standard model of biology (involving DNA, RNA and proteins) that has served us so well, and take into account the 'dark matter' of the biological universe — as well as the physical, biological, social and cultural environment.

**Ajit Varki** *University of California, San Diego, La Jolla, USA.*  
a1varki@ucsd.edu

## Getting the word out on biosphere crisis

A landmark statement released last week, formulated from a review published in *Nature* last year, illustrates how effectively and rapidly bridges can be built between science and society.

The review suggested that human influence may be forcing the global ecosystem towards a rapid, irreversible, planetary-scale shift (A. D. Barnosky *et al.* *Nature* **486**, 52–58; 2012). California's governor, Jerry Brown, invited

the authors to draw their findings to the attention of policy-makers, industry and the public. An international group of 16 scientists spent the next few months developing the statement and circulating it to the global-change scientific community.

The resulting statement — 'Maintaining humanity's life support systems in the 21st century' (see [go.nature.com/prudoq](http://go.nature.com/prudoq)) — is endorsed by more than 500 global-change researchers whose work spans every continent. It warns that unless decisive countermeasures are put into place immediately, climate change, loss of ecological diversity, extinctions, environmental contamination, human population growth and overconsumption of resources will degrade our quality of life within a few decades.

The statement is already catalysing interactions between governments, universities, business leaders and scientists to educate local constituencies about global issues and to stimulate solutions.

**Elizabeth A. Hadly\*** *Stanford University, California, USA.*  
hadly@stanford.edu

*\*On behalf of 4 co-signatories. See [go.nature.com/sgrxwo](http://go.nature.com/sgrxwo) for full list.*

## Boost to translational medicine in Europe

Next week will mark the launch of the European Infrastructure for Translational Medicine (EATRIS; see [go.nature.com/3li8fs](http://go.nature.com/3li8fs)). EATRIS aims to help clients (industry, small enterprises, funding organizations and academic institutions) to translate their biomedical discoveries efficiently into preventive, diagnostic or therapeutic products up to the point of clinical proof of concept.

More than 60 leading translational-research centres will make their facilities and expertise available to researchers and industry across Europe, through a single point of entry, bringing

together basic scientists and clinicians, as well as public and private partners. These centres may be specialists in, for example, advanced therapies, imaging or biomarkers.

EATRIS will minimize development costs and risks by matching expertise and optimizing the use of academic know-how, infrastructure and patient groups. Quality control, intellectual property and legal frameworks will all be standardized. Rapidly assembled multidisciplinary teams will address each project's specific regulatory and clinical needs.

These measures should also attract much-needed attention to rare and orphan diseases.

**Giovanni Migliaccio,**  
**Frank H. de Man, Anton E. Ussi**  
*EATRIS Coordination and Support, Amsterdam, the Netherlands.*  
giovannimigliaccio@eatris.eu

## Don't oversimplify psychiatric disorders

Psychiatric diagnosis is indeed imperfect, subjective and not based on pathophysiology or causation (*Nature* **496**, 416–418; 2013). But quests for biological markers and dimensional approaches, such as the Research Domain Criteria project mentioned, are unlikely to be clinically appropriate solutions because they run the risk of oversimplifying complex psychiatric illnesses and introducing predicative logic.

For instance, people with post-traumatic stress disorder (PTSD) or psychosis may show similar abnormalities of frontal-lobe function on a magnetic resonance imaging scan, just as those with anxiety or psychosis may show similar overactivation of the amygdala in the brain. But the underlying reasons for these superficial similarities are different, and so are the treatments. For example, lithium is not an effective treatment for psychosis, anxiety or PTSD, but it

is near-curative for some people who meet current diagnostic criteria for bipolar disorder.

The genetics of some conditions may overlap (for example, schizophrenia, autism and bipolar disorder), as do the genetics of, say, multiple sclerosis and Crohn's disease. However, these are distinct clinical conditions that require different interventions.

Caution and healthy scepticism are essential before embracing fashionable trends to revise psychiatric diagnosis.

**Daniel R. Weinberger** *Johns Hopkins University School of Medicine, Baltimore, Maryland, USA.*  
drweinberger@libd.org

## Communications need NIH funding

The US Congress launched an enquiry into all 'public relations' expenditure by the US National Institutes of Health (NIH) following coverage in *The Cancer Letter* of spending by the National Cancer Institute (NCI) Office of Communications and Education (OCE; see *Nature* **495**, 142; 2013). In our view, the criticisms devalue OCE activities and risk diverting attention away from the budget cuts that reduced the NIH's capacity to fund biomedical research.

Critics should note that several OCE activities are mandated or requested by Congress, and that OCE expenditure for the last fiscal year is less than 1% of the NCI's, so cutting this further will barely affect the NCI's research-funding capacity.

Communications are an important part of the NIH's mission. The NIH can only benefit from making NCI services and discoveries more accessible to cancer patients, physicians, researchers and the public.

**Judith S. Bond, Bethany Drehman** *Federation of American Societies for Experimental Biology, Bethesda, Maryland, USA.*  
bdrehman@faseb.org



## Time, space and memory

The brain's hippocampus contains place cells, which encode an animal's specific location. The finding that hippocampal neurons may also respond to time could provide information on the coding of episodic memories.

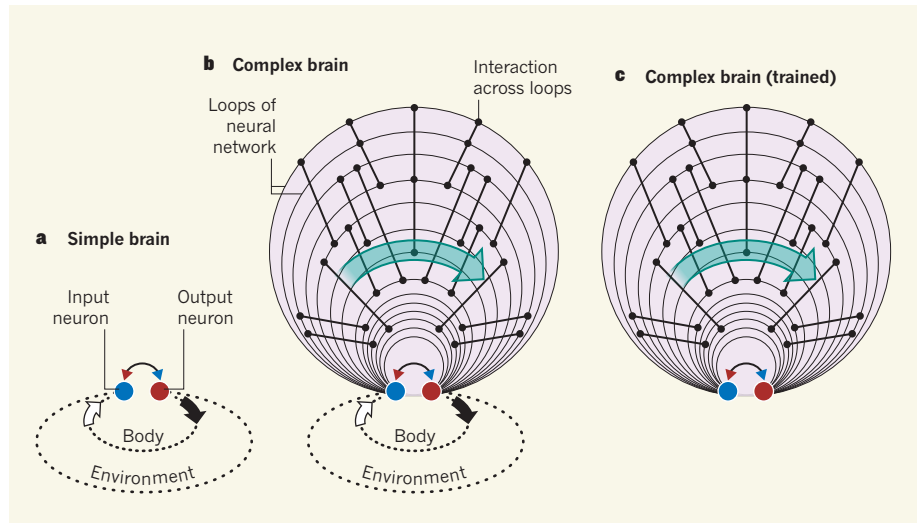
GYÖRGY BUZSÁKI

The philosopher Immanuel Kant made life difficult for neuroscientists by posing the following dilemma: does the brain represent time and space, which are themselves actual entities, or does it produce time and space and impose these categories on the world with which we interact? A more mundane, yet important, practical issue is whether there is a dedicated time-keeping mechanism in the brain, similar to the clock of a computer. The hippocampal region of the brain has long been suspected to represent Kantian space<sup>1</sup>. Now, in work published in *Neuron*, Kraus and colleagues<sup>2</sup> attempt to address the hard problem of time by describing 'time cells' — guess where, in the hippocampus. The implications of these findings are potentially far reaching.

The authors trained thirsty rats to run on a treadmill for tens of seconds for a reward of water, while recording the activity of groups of pyramidal cells in the animals' hippocampi. The aim was to distinguish neurons that 'track' elapsed time from those that track the distance run on the treadmill. Distance is, of course, a simple product of time and running speed. But the researchers manipulated these three variables by changing the speed of the treadmill from trial to trial, and requiring the rats to run for either a constant distance or a constant time on alternate days.

Members of the hippocampal-cell population that Kraus *et al.* recorded were active transiently and sequentially, so that the entire duration of the run was evenly represented by neuronal activity within this population. Using exemplary statistical and computational-modelling methods, the authors then evaluated the contribution of elapsed time and the distance run to the activity patterns of each recorded neuron.

In agreement with previous studies<sup>3,4</sup>, most of the recorded cells responded to an inseparable combination of time and distance. However, a minority of cells (still members of the broad distribution in the time–distance dimension) was mainly under the control of time spent on the treadmill, and the activity of an equally small fraction was distinctly correlated with distance. Kraus and co-authors



**Figure 1 | Externally driven and self-organized cell assemblies track time.** **a**, Evolutionarily simple brains contain simple neural networks. Sensory input from the body and the environment activates input neurons, which interact with output neurons to generate appropriate reflex actions in a short time window. **b**, In more complex brains, multiple interacting loops of increasing length improve prediction of more elaborate events that occur at longer timescales. **c**, After extensive training, the loops can sustain self-organized, long-lasting neuronal sequences without reliance on external cues and can, therefore, support cognitive operations such as memory, planning and imagination. Progression of neuronal operations correlates with elapsed time (green arrows) irrespective of whether the operations are driven externally or internally.

acknowledge that “it is impossible to completely separate time and distance”. Nonetheless, they interpret their results in favour of a dedicated mechanism in which this small population of ‘time cells’ specifically keeps track of time. This mechanism is distinct from that of path integration, by which distances and directions the animal takes are integrated with the help of the evolving cell assemblies.

The activity of time cells and path integration during navigation are only part of the hippocampus's story. This structure is also our resident ‘search engine’, which allows us to navigate in ‘mental space’ when recalling memories or planning future actions<sup>5</sup>. Kraus *et al.* suggest that the time cells they have identified are a key missing piece of episodic memory, the long-term memory that enables us to recall specific events and experiences, because such memories are embedded in a spatio-temporal context. To scrutinize this interpretation, it is useful to consider the operations

of the brain in a wider context.

Brains are predictive devices and exploit the fact that recurrence is a fundamental property of the world around us. Experience and memory allow the recall of similar situations and the deployment of previously effective actions. In simple neuronal circuits, such as those of invertebrates, signals from the environment or the body can trigger appropriate (learned) responses within a relatively short time window (Fig. 1). With increasing organismal complexity, ever-increasing loops of neuronal networks are added to the basic circuit to improve prediction of more complex events and those with longer temporal separation between the input signals and the responses. After sufficient training, the long loops of larger brains can dispense with the reliance on external cues by internally processing the probabilities of outside events and their most likely outcomes. This disengagement is a necessary condition for cognition<sup>6</sup>.

Returning to the findings of Kraus and colleagues, these suggest that progression of neuronal information within cell assemblies in the hippocampus during spatial navigation can be controlled by environmental or body cues (for path integration) and by a time-tracking mechanism. Alternatively, episode-specific activity sequences of cell assemblies may roll forward as a result of self-organization in the absence of changing external cues, always progressing along the path of the highest-probability events<sup>3</sup>. This latter perspective suggests that the qualitative distinction between the causes of sequential neuronal activity is along the dimension of external dependence versus internal self-organization, rather than being related to time or distance.

Brains, like clocks, do not produce time per se. So despite our intuition of separate time and space, it may be that the brain generates no such things. Instead, variations in the strength of synaptic communication between neurons may simply determine the direction of activity flow across neurons under all conditions. The temporal flow of activity is a framework for recalling thousands of episodic memories or planning multiple possible consequences of actions. The evolving neuronal-assembly sequences that support these cognitive operations may activate all hippocampal neurons at some point, including those that may sometimes appear in the disguise of 'time-tracking-only' cells.

Many brain regions can generate sequential activity paced for their own needs<sup>3,7-9</sup> — ranging from the subsecond scale in the service of perception and motor control to much longer timescales in memory, planning and imagination. These operations progress along a time line but do not need appointed time cells. Admittedly, the questions Kraus and colleagues address are among the most complex in science, and the authors should be commended for tackling a hard problem at the intersection of philosophy and neuroscience. As is always the case with good science, their findings raise as many questions as they tried to answer. I have addressed only a few, and leave the rest to the Kant scholars. ■

**György Buzsáki** is at the Neuroscience Institute, New York University Langone Medical Center, New York, New York 10016, USA.  
e-mail: gyorgy.buzsaki@nyumc.org

1. O'Keefe, J. & Nadel, L. *The Hippocampus as a Cognitive Map* (Oxford Univ. Press, 1978).
2. Kraus, B. J., Robinson, R. J. II, White, J. A., Eichenbaum, H. & Hasselmo, M. E. *Neuron* <http://dx.doi.org/10.1016/j.neuron.2013.04.015> (2013).
3. Pastalkova, E., Itskov, V., Amarasingham, A. & Buzsáki, G. *Science* **321**, 1322–1327 (2008).
4. Itskov, V., Curto, C., Pastalkova, E. & Buzsáki, G. *J. Neurosci.* **31**, 2828–2834 (2011).
5. Tulving, E., Donaldson, W. & Bower, G. H. (eds)

*Organization of Memory* (Academic, 1972).

6. Buzsáki, G. *Rhythms of the Brain* (Oxford Univ. Press, 2006).
7. Mauk, M. D. & Buonomano, D. V. *Annu. Rev. Neurosci.* **27**, 307–340 (2004).

8. Fujisawa, S., Amarasingham, A., Harrison, M. T. & Buzsáki, G. *Nature Neurosci.* **11**, 823–833 (2008).
9. Harvey, C. D., Coen, P. & Tank, D. W. *Nature* **484**, 62–68 (2012).

## GENOMICS

# A spruce sequence

**The first published whole-genome draft sequence of a gymnosperm, the Norway spruce, provides a powerful platform for studying the unique development, adaptation and evolution of this major group of plants. SEE ARTICLE P.579**

RONALD SEDEROFF

**W**ithin the gymnosperm subgroup of seed plants are the iconic conifers, which dominate many forest ecosystems of the cold-temperate and subtropical regions of the Northern Hemisphere. Among the conifers are Earth's oldest living individual plants, the bristlecone pines; its largest trees, the giant sequoias; and its tallest, the coast redwoods. Conifers also include the pine and spruce genera that supply much of the world's wood for pulp, paper and solid-wood products. Genetic analysis is the key to understanding the biology of these trees, but gymnosperms typically have very large genomes, of up to 37 gigabases<sup>1</sup>, and an abundance of repetitive DNA, making their sequences difficult to assemble. So the sequence of the Norway spruce genome, reported by Nystedt *et al.*<sup>2</sup> on page 579 of this issue, represents a major technical, as well as an information-rich, achievement\*.

Gymnosperms, of which there are around 1,026 species<sup>3</sup>, are vascular plants, meaning that they contain a tubular tissue network that transports water and nutrients, and provides mechanical support. The gymnosperms are one of two subgroups of seed-forming vascular plants, the other being the angiosperms, the flowering plants. There are some 350,000 species of angiosperm, some woody and others herbaceous, and this subgroup includes all our major food crops.

The seeds of angiosperms are enclosed in an ovary, whereas those of gymnosperms are in an open state. The two subgroups also differ in terms of their mechanisms of growth, development, metabolism, adaptation and evolution<sup>4</sup> — in factors including wood microanatomy, water transport, mechanical support, reproduction, development and ability to adapt to environmental change. The vascular system of conifers depends on long, thin cells with lateral pits called tracheids, a primitive system also found in early progymnosperm fossils. Angiosperms typically have

\*This article and the paper under discussion<sup>2</sup> were published online on 22 May 2013.

more complex wood-cell anatomy, in which vessels with much larger diameters facilitate water transport and more-specialized fibre cells give mechanical support. The phenolic polymer lignin contributes to mechanical support and provides a hydrophobic surface for water transport in both angiosperms and gymnosperms, but the composition of the polymer differs markedly between the two groups of plants.

Identifying the genes that underlie these differences is of interest for both basic and applied research, but the long reproductive



**Figure 1 | Gymnosperm genes.** Trees of the gymnosperm subgroup make up much of the forests of the Northern Hemisphere and provide a large fraction of the world's wood. Nystedt *et al.*<sup>2</sup> have presented the first draft whole-genome sequence of a gymnosperm, the Norway spruce (*Picea abies*).

JOSEPH STRAUCH/GETTY



cycles and large sizes of gymnosperms have made traditional, breeding-based analyses of these plants challenging. DNA-based technology that can bypass these limitations has been particularly useful in forest trees, enabling genomic mapping, gene sequencing, genomic selection and genetic engineering. Whole-genome sequences are particularly powerful, because they provide a platform for a spectrum of new technologies, such as studies of an organism's full transcriptome — its complement of transcribed RNA molecules. Nystedt and colleagues' draft sequence is the first gymnosperm sequence to be published (Fig. 1), and a loblolly pine genome sequence (around 22 Gb) is expected to follow soon<sup>5</sup>. These conifer genomes are the largest plant genomes sequenced so far, much larger than the 17-kb genome of wheat<sup>6</sup> (an angiosperm).

The spruce genome will not only accelerate the investigation of gymnosperm biology, it will also provide broader genetic and evolutionary insight. For example, researchers of the ENCODE project<sup>7</sup> recently argued that 70% of the 3.2-Gb human genome is functional in some way. But large plant and animal genomes pose a challenge to this proposal. Consider, for instance, the spruce genome (around 20 Gb) and that of thale cress, the model plant *Arabidopsis thaliana* (0.135 Gb). If a similarly high proportion of these genomes were functional, what properties of gymnosperms could necessitate such a massively larger number of functional genetic elements? Repression of only two genes enables *A. thaliana* to extend its growth cycle and produce a substantial amount of wood<sup>8</sup>, suggesting that the large genomes of gymnosperms are not attributable to a perennial growth habit or wood formation. Early comparisons of transcribed genes indicated<sup>9</sup> that the vast majority of genes in gymnosperms and angiosperms have homologues in the same gene families. Nystedt and colleagues estimate that the spruce genome contains around 28,354 genes, which is very close to *A. thaliana*'s 27,407. The angiosperms maize (corn), rice and poplar have estimated gene numbers of around 40,000. But these coding regions comprise a fraction of the sequence of these large genomes, and what the functions of the remaining sequences may be is still obscure.

Another curious aspect of gymnosperm genomes is the evolutionary conservation, in many species, of a haploid (single copy) chromosome number of 12, despite their genome sizes ranging from 9.7 to 37 Gb<sup>1</sup>. The genomes will also provide insight into mechanisms of ancient and recent evolutionary adaptation in plants. Gymnosperms are thought to have originated from progymnosperms 360 million years ago, but much about the origin of angiosperms remains a mystery. Although the fossil record supports a gymnosperm origin for angiosperms, estimates of the time

of angiosperm origins vary by more than 100 million years. There is also controversy regarding relationships within the gymnosperms, particularly the relationship of the conifers to the gnetophytes, which include the bizarre desert plant *Welwitschia*.

The genome sequences of these trees will not only help us to understand the past, but may also increase our understanding of present-day northern-latitude forests. Gymnosperms became the dominant forest plants in the late Palaeozoic and the Mesozoic periods, around 300 million to 70 million years ago. But during the most recent ice age, much of the northern latitudes were covered by ice. When the glaciers last retreated, only about 10,000 years ago, the conifers were the major pioneer species that dominated that land. Understanding how the gymnosperms established new ecosystems as the glaciers shrank is becoming more important as we anticipate the

effects of global climate change on the world's forests. ■

**Ronald Sederoff** is in the Forest Biotechnology Group, Department of Forestry and Environmental Resources, North Carolina State University, Raleigh, North Carolina 27695, USA.  
e-mail: ron\_sederoff@ncsu.edu

1. Ahuja, M. R. & Neale D. B. *Silvae Genetica* **54**, 126–137 (2005).
2. Nystedt, B. *et al. Nature* **497**, 579–584 (2013).
3. Christenhusz, M. J. M. *et al. Phytotaxa* **19**, 55–70 (2011).
4. Beck, C. B. *Origin and Evolution of Gymnosperms* (Columbia Univ. Press, 1988).
5. Pine Reference Sequences [pinegenome.org/pinerefseq](http://pinegenome.org/pinerefseq)
6. Brenchley, R. *et al. Nature* **491**, 705–710 (2012).
7. Dunham, I. *et al. Nature* **489**, 57–74 (2012).
8. Melzer, S. *et al. Nature Genet.* **40**, 1489–1492 (2008).
9. Kirst, M. *et al. Proc. Natl Acad. Sci. USA* **100**, 7383–7388 (2003).

## PLANETARY SCIENCE

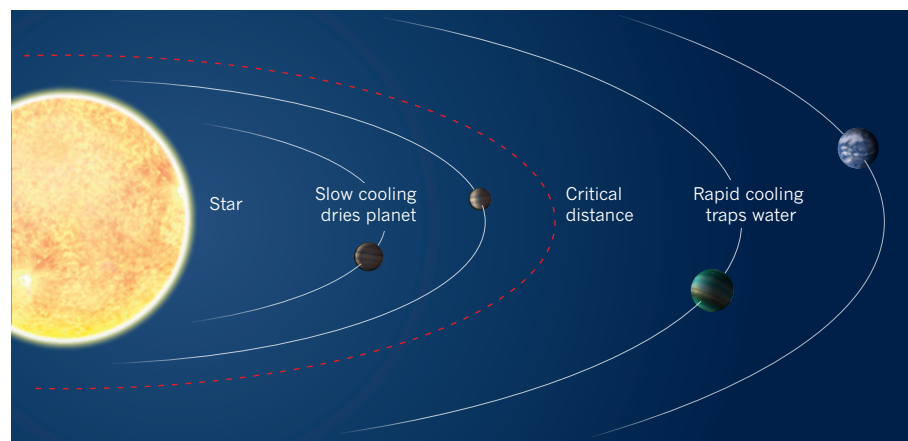
## Evolutionary dichotomy for rocky planets

**A simple model shows that a rocky planet close to its star may solidify so slowly that its water is lost to space and the planet becomes desiccated, whereas a planet farther out may solidify quickly and retain its water. SEE LETTER P.607**

LINDA T. ELKINS-TANTON

Earth and Venus were probably built from similar rocky materials, having been formed by similar mass-accretion processes. The giant impacts that characterize these processes are thought to melt the

growing planets to some depth, producing one or more magma-ocean stages during which the silicate portion of the planet is melted before solidifying. Thus, there has been no reason to suspect that Venus and Earth differed through the first tens and probably hundreds of millions of years of the Solar System, and Venus



**Figure 1** | Hamano and colleagues' evolutionary model for terrestrial planets<sup>1</sup>. Rocky planets that form and solidify close to their host stars may receive enough heat from the star to slow cooling, allowing time for water to escape into space and causing the planet to dry out. Rocky planets that form farther from their stars, beyond a critical distance, would cool quickly, trapping water in silicate minerals in their interiors and as liquid on their surfaces.

is commonly thought to have lost its water through some later divergence from Earth-like evolution. On page 607 of this issue, Hamano *et al.*<sup>1</sup> present a simple model that might explain why rocky planets that have similar compositions but orbit at different distances from their host stars can end their magma-ocean stages with either an Earth-like wetness or a Venus-like dryness. This model does not require any later divergence to explain the differences between the planets.

Almost 30 years ago, researchers showed how a dense steam atmosphere can be generated on a young, hot planet by the solidification of an impact-generated magma ocean<sup>2</sup>. After the upper troposphere (the lowest portion of the atmosphere) of the young planet has become saturated with steam, that atmospheric layer imposes a strict upper limit on outgoing radiation from the magma ocean — about 300 watts per square metre. Therefore, as soon as the magma ocean produces a steam-saturated troposphere, the cooling rate of the planet is controlled by this one simple limit.

Previously, several groups had calculated that a magma ocean should solidify in just millions of years<sup>3–5</sup>. These calculations assumed that the planet had lower incoming heat flux from the star than outgoing heat flux from the magma ocean. The crucial feature of Hamano and colleagues' model is that some planets are close enough to their star for the incoming heat flux to be higher than the  $300 \text{ W m}^{-2}$  outgoing radiation limit, and thus the planet would be prevented from cooling at all until water was lost from the steam-saturated atmosphere.

For planets close to their star, solidification and cooling together may take orders of magnitude longer than for more distant planets, perhaps as long as hundreds of millions of years. After solidification, cooling proceeds only as water is stripped from the hot, inflated atmosphere by hydrodynamic escape. The longer that solidification and cooling take, the more water is lost to space, and the drier the planet becomes. Thus, the distance of a terrestrial planet from its host star might produce an evolutionary dichotomy (Fig. 1).

The authors further suggest that Earth solidified far enough from the Sun to have a net loss of planetary heat from the beginning, allowing it to solidify quickly. Earth's initial water inventory influenced the volume of only its initial oceans. Venus, however, may have had net heat flux into the planet, and its current dryness might be related to this early slow solidification and attendant atmospheric water loss, before cooling allowed the water in the steam atmosphere to cool and condense into liquid oceans.

Recent work on geochemical tracers has indicated that Earth's mantle solidified and differentiated from a magma ocean more than 4.45 billion years ago<sup>6</sup>, probably around 4.52 billion years ago<sup>7</sup>, which agrees with rapid

solidification. For Venus, however, there are insufficient geochemical data to perform this test. Measurements of deuterium and hydrogen in the Venusian atmosphere indicate that the planet has lost a substantial amount of water over time<sup>8,9</sup>, but whether that loss occurred at the time of solidification or more recently is a matter of argument.

The authors' model underscores the importance of the earliest accretion and solidification steps in determining the future evolution of the rocky planets. However, several crucial caveats need to be considered in applying this model. First, in extrapolating back in time, the faint young star's radiation level needs to be considered. Second, initial atmospheres might not all be water-rich; the rocky building blocks for some planets might have produced atmospheres rich in methane and hydrogen, instead of steam<sup>10</sup>. In the absence of a steam atmosphere, there would be no outgoing radiation limit to slow solidification and cooling. Third, forming an initial atmosphere above a magma ocean is not a simple process. The removal of volatile gases from magma might require a significant degree of supersaturation and might not occur until late in solidification. If this is so, then solidification would proceed to a high degree before a steam atmosphere formed and occluded heat flux.

Although proximity to a star affects planetary water content, this is not the only parameter that dictates the habitability of a rocky planet — the planet's composition also has a strong influence on all aspects of habitability, such as bulk atmospheric composition, susceptibility to plate tectonics and formation of a shielding magnetic field. A challenge for the coming decades will be to make measurements of exoplanets that allow the testing of models for habitability, and these tests need to include composition. How do atmospheric species other than water affect the solidification rates of magma oceans? What atmospheric compositions would be expected in the wake of a slow solidification with substantial water loss? The habitability of Earth and the inhospitability of Venus may be the inevitable result of our planetary sibling order next to the Sun rather than later evolutionary bifurcations. If so, similar patterns of habitability are likely to be found in exoplanets. ■

**Linda T. Elkins-Tanton** is in the Department of Terrestrial Magnetism, Carnegie Institution for Science, Washington DC 20015, USA.  
e-mail: ltelkins@dtm.ciw.edu

1. Hamano, K., Abe, Y. & Genda, H. *Nature* **497**, 607–610 (2013).
2. Abe, Y. & Matsui, T. *J. Geophys. Res.* **90**, C545–C559 (1985).
3. Abe, Y. *Phys. Earth Planet. Int.* **100**, 27–39 (1997).
4. Zahnle, K. J., Kasting, J. F. & Pollack, J. B. *Icarus* **74**, 62–97 (1988).
5. Elkins-Tanton, L. T. *Earth Planet. Sci. Lett.* **271**, 181–191 (2008).



6. Mukhopadhyay, S. *Nature* **486**, 101–104 (2012).
7. Touboul, M., Puchtel, I. S. & Walker, R. J. *Science* **335**, 1065–1069 (2012).
8. Donahue, T. M., Hoffman, J. H., Hodges, R. R. Jr & Watson, A. J. *Science* **216**, 630–633 (1982).
9. Zahnle, K. J. & Kasting, J. F. *Icarus* **68**, 462–480 (1986).
10. Hashimoto, G. L., Abe, Y. & Sugita, S. J. *Geophys. Res.* **112**, E05010 (2007).

## SYNTHETIC BIOLOGY

# It's an analog world

**The first synthetic genetic circuits to use analog computation have been developed. These circuits involve fewer components and resources, and can execute more complex operations, than their digital counterparts. SEE LETTER P.619**

HERBERT M. SAURO & KYUNG HYUK KIM

The engineering of biological networks, a discipline called synthetic biology, has seen remarkable progress since its inception in 2000. The fact that we can now design simple but functional circuits *in vivo* is probably one of the most important aspects of this progress — it suggests that our biophysical understanding of the cellular milieu is largely correct, even if many details remain to be resolved. The significance of this should not be underestimated, because it means that predictive engineering of new cellular networks and systems is possible. Until now, approaches to synthetic biology have been greatly influenced by the digital computing technology that we use every day. But biological cells do not process information in a solely digital fashion; rather, they carry out many operations in an analog manner. In this issue, Daniel *et al.*<sup>1</sup> (page 619) break the mould and venture into a completely new area of synthetic biology, presenting synthetic circuitry that can perform analog calculations\*.

Synthetic biology has already led to several notable firsts — from the construction of switches, oscillators and feed-forward networks to the creation of gene regulatory circuits that can carry out Boolean logic. In many ways, digital computing has coloured our approach to synthetic biology. It is a truly remarkable achievement that so much can be done with just the 'ones' and 'zeros' that form the basis of digital processing. After more than 60 years of digital computing, it is no surprise that many modes of thinking have become highly influenced by the digital paradigm. It is often said that a biological cell is just like a digital computer, and that reprogramming a cell is like writing software. But the digital analogy is a misleading one, because many processes that occur in a biological cell have no counterpart in a digital computer. If reprogramming a cell was just a case of writing code, then re-engineering cells would be simple — but it is not.

\*This article and the paper under discussion<sup>1</sup> were published online on 15 May 2013.

The reality is that cells use a hybrid approach to information processing. In some cases they use digital yes-or-no decisions, but in many cases cellular signals are analog, with levels of gradation. More exotic, little understood signal-processing techniques involving noise and other forms of signal probably also contribute. And on top of that, a complex chemistry exists that continuously reassembles the cell in real time.

Nevertheless, constructing digital devices is both a useful and an interesting engineering challenge for synthetic biologists, and many advances have been made in mimicking digital

**“The ability to process graded information by means of synthetic biological circuits will be of interest to many researchers.”**

systems either *in vivo* or *in vitro* using DNA as components. One remarkable study was the construction of an *in vitro* 4-bit square-root calculator consisting of 130 DNA strands<sup>2</sup>. Another was the construction of an *in vivo* system to detect four

different inputs, comprised of four sensors and three logic gates<sup>3</sup>. Although these are significant achievements, such circuits require large numbers of components to perform even the simplest computation. A 4-bit binary adder, for example, might require 30 or more proteins to operate, and at the same time would place a substantial metabolic burden on a cell. More challenging and perhaps more interesting is to attempt to mimic analog computation in a cell, as Daniel *et al.* have done.

One key advantage of analog over digital is that far fewer devices are needed to carry out a given computation at the moderate precision needed in cells<sup>4</sup>, and fewer devices mean lower resource requirements. In addition, the richness of signal processing that can be carried out in analog systems is far greater than what can be accomplished in digital systems using the same number of components. But analog information processing brings with it a new set of challenges. Compared with that of

digital systems, the design of analog circuitry requires greater expertise in feedback-system design, circuit theory and signal processing. Daniel and colleagues' work is founded on the close analogy they saw between the exponential thermodynamic electron flow that occurs in transistors and the exponential thermodynamic rates seen in chemical reactions<sup>5</sup>. This similarity led them to realize that analog electronic circuits that operate in the logarithmic domain might be effective mimics of analog biological circuits that operate in the logarithmic domain.

To test this idea, Daniel *et al.* constructed a gene-transcription unit in which the concentration of a transcription factor (a protein that regulates gene transcription) could be logarithmically transformed over a wide range of concentrations, thus enabling fine control of gene expression.

Moving an input signal into the logarithmic domain has many benefits, especially that basic calculations such as division and multiplication can be more easily executed in log space. The core of the authors' logarithmic unit is a positive-feedback loop operating in an analog mode. A shunt is included that removes excess transcription factor so as to increase the range of operation and efficiently implement positive feedback. With this basic logarithmic unit in place, the researchers could then focus on adding higher-order functions. For example, by attaching a repressor module, they created a circuit whose readout was the negative of the log of the input. Another circuit simply summed the activity of two parallel logarithmic circuits to mimic a multiplication operation.

Furthermore, using just two transcription factors, the authors were able to create a negative-feedback analog circuit that scaled the logarithmic function. Especially interesting was their construction of a circuit that computed the ratio of two inputs over four orders of magnitude of input concentration. Computing a ratio might have many applications, such as normalizing or comparing values, or even providing an *in vivo* pH meter.

The ability to process graded information by means of synthetic biological circuits will be of interest to many researchers. For example, most inputs from the environment are graded, and a synthetic circuit with a fluorescent protein readout might be able to represent the rate of change of an environmental input rather than its absolute level. This could be accomplished by having an analog differentiator circuit. Other analog designs might be used to measure the weighted sum of environmental inputs, such that when a particular combination of inputs is reached it triggers a change in the state of the cell. Designing such analog circuitry may also further our understanding of natural systems. The function of many regulatory systems in cells remains obscure, and studies of synthetic biology in the analog domain should lead to

new theories on how biological systems process information, and thus allow such systems to be more finely controlled. ■

**Herbert M. Sauro and Kyung Hyuk Kim**  
are in the Department of Bioengineering,  
University of Washington, Seattle,

Washington 98195-5061, USA.  
e-mails: hsauro@u.washington.edu;  
kkim@uw.edu

1. Daniel, R., Rubens, J. R., Sarpeshkar, R. & Lu, T. K. *Nature* **497**, 619–623 (2013).
2. Qian, L. & Winfree, E. *Science* **332**, 1196–1201 (2011).

3. Moon, T. S., Lou, C., Tamsir, A., Stanton, B. C. & Voigt, C. A. *Nature* **491**, 249–253 (2012).
4. Sarpeshkar, R. *Neural Comput.* **10**, 1601–1638 (1998).
5. Sarpeshkar, R. *Ultra Low Power Bioelectronics: Fundamentals, Biomedical Applications, and Bio-Inspired Systems* (Cambridge Univ. Press, 2010).

## PALAEOANTHROPOLOGY

# Hesitation on hominin history

**Extensive studies of fossil skeletons of *Australopithecus sediba* provide fascinating details of the anatomy of this hominin species, but do not convincingly indicate its position on the evolutionary route to modern humans.**

WILLIAM H. KIMBEL

The evolutionary events that led to the origin of the *Homo* lineage are an enduring puzzle in palaeoanthropology, chiefly because the fossil record from between 3 million and 2 million years ago is frustratingly sparse, especially in eastern Africa. Much attention has been paid to two fossilized skeletons, found in approximately 2-million-year-old sediments at the Malapa cave site in South Africa, that are recognized as representing the species *Australopithecus sediba*. These have been the focus of scrutiny because of both their excellent preservation and claims<sup>1,2</sup> that this hominin — a species more closely related to humans than to chimpanzees — lies at the base of the *Homo* lineage. A series of reports published in *Science*<sup>3–8</sup> sheds light on the morphology of *A. sediba* but, in my view, does little to elucidate its role in later human evolution.

Dental morphology is a frequent source of information about hominin phylogeny but, in the first of these new papers, Irish *et al.*<sup>3</sup> take the unconventional step of using only the Arizona State University Dental Anthropology System — a graded series of minor crown variants originally devised to distinguish recent human populations from one another — to decipher relationships between hominin species that are millions of years old. I have serious doubts about the phylogenetic meaning of morphological similarity in this case. These concerns are compounded by the authors' reliance on the gorilla as the sole outgroup in their cladistic analysis. Their results link *A. sediba* exclusively to *Australopithecus africanus*, an older (approximately 2.7 million to

2.3 million years old), potentially ancestral, southern African species with which it also shares some key cranial features<sup>1</sup>. If this finding is borne out by further work, then the relevance of *A. sediba* to the origin of *Homo* would be inextricably tied to that of *A. africanus*, whose own position in hominin phylogeny is by no means settled<sup>9</sup>.

De Ruiter and colleagues' analysis of the *A. sediba* mandible<sup>4</sup> includes a measurement-based comparison in which the sub-adult individual MH1 (with only its second molar

erupted) is treated as though its growth had been completed. However, for most dimensions, hominoid mandibles achieve only around 75–90% of their adult values by the time of the second molar eruption<sup>10</sup>. So, although the *A. sediba* mandibles seem to be small and lightly built (and thus *Homo*-like) by australopithec standards, it is unclear how much of this impression is due to the authors' use of a sample comprising a sub-adult and a presumed adult female (MH2).

Much of the value of the Malapa material lies in the extremely rare association of upper and lower limb parts with elements of the axial skeleton in two individuals of the same species (Fig. 1). These skeletons paint a portrait of a pectoral girdle that retains more ape-like anatomy than the pelvic girdle<sup>5,11</sup>. Churchill *et al.*<sup>5</sup> report that a fairly complete scapula (from MH2) features an upwardly tilted articulation for the humerus and a relatively broad attachment area for a muscle that helps to lift the arm over the head, a familiar australopithec upper-limb pattern that also includes long, strong forearms and curved fingers. Although these features are embedded in a terrestrial bipedal

frame, they are often interpreted as signs of retained ancestral arboreal climbing behaviour<sup>5,12</sup>. Still unsettled is what led to the refashioning of the hominin shoulder by the time, around 1.6 million years ago, of *Homo erectus*, a species that shows modern upper-limb and shoulder morphology (this anatomy is unknown in the approximately contemporaneous *Homo habilis*). Simply leaving the trees seems to be an insufficient explanation.

Schmid *et al.*<sup>6</sup> used the low curvature of the upper ribs of *A. sediba* to argue for a conical ribcage and elevated shoulders similar to those of the great apes, even though second- and fourth-rib curvatures do not actually distinguish apes from humans. However, it is clear that the unusually strongly curved first rib articulates only with the first thoracic vertebra, as in humans and *Australopithecus afarensis*. This configuration is at odds with a completely ape-like upper thorax and has been associated with descent of the shoulder after the upper limbs were freed from locomotion<sup>13</sup>, although this interpretation has been contested<sup>14</sup>. A further puzzle is the *A. afarensis* partial skeleton KSD-VP 1/1, which, although 1.6 million years older than the *A. sediba* skeletons, has



**Figure 1 | *Australopithecus sediba*.** A series of papers<sup>3–8</sup> presents extensive studies of these two fossil skeletons, which date to approximately 2 million years ago. The authors compare the anatomy of this hominin to that of other species of the *Australopithecus* and *Homo* genera.

BRENT STIRTON, COURTESY LEE R. BERGER & UNIV. WITWATERSRAND



an upper thorax more similar to that of modern humans<sup>15</sup>. A lower (ninth) rib from MH2, which is more like those of modern humans in its curvature and torsion, is consistent with the less flaring pelvic rim of *A. sediba* when compared with *A. afarensis* and *A. africanus*<sup>6,11</sup>.

The 'long-backed' lumbosacral vertebral formula of six lumbar and four sacral vertebrae that is seen in species from australopiths through to *Homo erectus* is the probable primitive condition for humans and the great apes<sup>16</sup>. The evolution of the most common modern-human condition of five lumbar and five sacral vertebrae occurred by 'sacralization' of the lowermost lumbar vertebra. Williams *et al.*<sup>7</sup> argue that MH2 is unusual for an early hominin in having this derived pattern, but the MH2 formula depends on how one defines a lumbar vertebra — by a lack of rib articulations or by functional criteria that relate to intervertebral movement. In early hominins, the first of the six functional lumbar vertebrae carries rib articulations that are similar to those of a thoracic vertebra. MH2, in fact, resembles other *Australopithecus* specimens in having six functional lumbar, but this is unexpected in an early hominin with five sacral vertebrae.

The last of the papers presents DeSilva and colleagues' reconstruction of the *A. sediba* gait<sup>8</sup> (based on the MH2 skeleton), which will be controversial. The proposed 'hyperpronation' of the foot and extreme inward rotation of the leg and thigh suggest an ungainly bipedal stride that might have made it into Monty Python's 'Ministry of Silly Walks' sketch. The presumed inversion of the foot at the heel-strike of the surprisingly ape-like calcaneus, combined with a vertical shank (tibia), outwardly angled thigh (femur) and a long, lordotic lower back — all hallmarks of terrestrial bipedality in *Australopithecus* and *Homo* species — constrains the reconstruction. Prominent osteophytic growths on the pelvis and fibula at the attachment sites of the thigh musculature raise the possibility of a gait that was pathologically impaired, but DeSilva and colleagues argue that this locomotor pattern was adaptive. However, if *A. sediba* was a descendant of *A. africanus*, which, similarly to the even older *A. afarensis* (dating to between 3.7 million and 3.0 million years ago), shows no trace of this pattern, then it is hard to imagine the selective advantage that would accrue from such a kinematically peculiar gait.

Given the mix of features seen in *A. sediba*, it is difficult to understand why these researchers insist that it lies at the base of the *Homo* lineage<sup>1,2</sup>. Similar intellectual gymnastics are required to comprehend the authors' argument that no African *Homo* fossils exist from before the time of *A. sediba*<sup>2,17</sup>. Although the recent papers constitute a fascinating further analysis of the *A. sediba* fossils, I do not think that they provide compelling evidence that this species is anything other than an unusual

australopith from a Pliocene–Pleistocene time period that is already populated by a fair number of them<sup>18</sup>. ■

**William H. Kimbel** is at the Institute of Human Origins and the School of Human Evolution and Social Change, Arizona State University, Tempe, Arizona 85287-4101, USA. e-mail: [wkimbel.ih@asu.edu](mailto:wkimbel.ih@asu.edu)

- Berger, L. R. *et al.* *Science* **328**, 195–204 (2010).
- Pickering, R. *et al.* *Science* **333**, 1421–1423 (2011).
- Irish, J. D., Guatelli-Steinberg, D., Legge, S. S., de Ruiter, D. J. & Berger, L. R. *Science* <http://dx.doi.org/10.1126/science.1233062> (2013).
- de Ruiter, D. J. *et al.* *Science* <http://dx.doi.org/10.1126/science.1232997> (2013).
- Churchill, S. E. *et al.* *Science* <http://dx.doi.org/10.1126/science.1233477> (2013).
- Schmid, P. *et al.* *Science* <http://dx.doi.org/10.1126/science.1234598> (2013).
- Williams, S. A. *et al.* *Science* <http://dx.doi.org/10.1126/science.1232996> (2013).

- DeSilva, J. M. *et al.* *Science* <http://dx.doi.org/10.1126/science.1232999> (2013).
- Grine, F. E. in *The Paleobiology of Australopithecus* (eds Reed, K. E., Fleagle, J. G. & Leakey, R. E.) 73–104 (Springer, 2013).
- Humphrey, L. T. in *Human Growth in the Past: Studies from Bones and Teeth* (eds Hoppa, R. & FitzGerald, C.) 65–87 (Cambridge Univ. Press, 1999).
- Kibii, J. M. *et al.* *Science* **333**, 1407–1411 (2011).
- Larson, S. G. in *The Paleobiology of Australopithecus* (eds Reed, K. E., Fleagle, J. G. & Leakey, R. E.) 247–262 (Springer, 2013).
- Ohman, J. C. *Am. J. Phys. Anthropol.* **70**, 209–229 (1986).
- Stern, J. T. Jr & Jungers, W. L. *Am. J. Phys. Anthropol.* **82**, 431–439 (1990).
- Haile-Selassie, Y. *et al.* *Proc. Natl Acad. Sci. USA* **107**, 12121–12126 (2010).
- McCollum, M. A., Rosenman, B. A., Suwa, G., Meindl, R. S. & Lovejoy, C. O. *J. Exp. Zool.* **314B**, 123–134 (2010).
- Berger, L. R. *J. Anthropol. Sci.* **90**, 1–16 (2012).
- Kimbel, W. H. in *Handbook of Paleoanthropology* Vol. 3 (eds Henke, W. & Tattersall, I.) 1539–1574 (Springer, 2007).

#### ASTROPHYSICS

## A glimpse inside a magnetar

**Hundreds of neutron stars have exhibited 'glitches' in their spin-down rates — an indication of ultra-dense superfluids in their interiors. Now one highly magnetized star has shown a surprising glitch in the 'wrong' direction. SEE LETTER P.591**

ROBERT C. DUNCAN

A neutron star resembles a giant atomic nucleus, with 1–2 times the Sun's mass packed into a ball about 20 kilometres across. Its gravity is so strong that a projectile would need to be launched at about half the speed of light to escape from its surface. Extreme density, pressure, temperature, magnetism and relativistic gravity make these objects fascinating but challenging to study. Surprising observations of spin-down irregularities in one intensely magnetized neutron star, reported by Archibald *et al.*<sup>1</sup> on page 591 of this issue, offer clues about exotic processes occurring deep inside these objects.

The basic structure of a neutron star is generally agreed on. It has a crust about 1 km thick, in which nuclei are arranged in a crystal lattice immersed in a 'sea' of electrons. Near the surface, the nuclei are plain iron, but the pressure and density increase rapidly with depth, so that the nuclei become increasingly bloated and neutron-rich. At moderate depth, neutrons 'drip' out of the nuclei, forming a neutral liquid between the lattice nuclei. At the base of the crust, the bloated nuclei merge. Below this lies pure nuclear fluid, more than 200 trillion times denser than liquid water.

Unlike atomic nuclei, which contain almost equal numbers of protons and neutrons, the nuclear fluid inside a neutron star has roughly 20 neutrons for every proton, a ratio maintained by neutrino-emitting processes. There is also one electron or muon ('heavy electron') per proton, ensuring charge neutrality. These particles are forced together by tremendous pressure, but quantum mechanics requires them to occupy different states, so they fill all available energy states up to a high energy, the Fermi energy. Deep below the crust, as the pressure rises, neutron and proton Fermi energies get so high that exotic, strongly interacting particles such as hyperons and mesons might join the mix. The actual fluid composition at very high densities is uncertain. It is possible that the innermost, central core consists of a 'soup' of quarks, the elementary particles that make up protons and neutrons.

Fortunately, observations of neutron stars yield insight into their interiors. The most thoroughly studied neutron stars are radio pulsars, which emit radio blips as they rotate. Timing the blips reveals that these stars steadily spin down. This is due to their intrinsic magnetism: as radio pulsars spin, they blow out magnetic waves and winds of fast charged particles, which carry away angular momentum. Interestingly, this steady spin-down is

occasionally punctuated by ‘glitches’ — incidents in which the spin rate abruptly ‘jumps up’ by a small fractional amount. Hundreds of radio pulsars have been accurately timed, and many hundreds of glitches observed<sup>2,3</sup>. The glitches are attributed to imperfect coupling of superfluids within the stars: as a neutron star spins down, superfluid components tend to spin faster than the rest of the star. A glitch occurs when the superfluid occasionally shifts closer to co-rotation.

Although young neutron stars are hot, even by astrophysical standards, with interior temperatures commonly on the order of  $10^8$  kelvin, their strongly interacting neutrons can pair up and rearrange into quantum superfluids: phenomena found in low-temperature physics laboratories<sup>4</sup>. In the inner crust, below the level of ‘neutron drip’, the pairing neutrons have zero relative angular momentum, analogous to Cooper electron pairs in a superconductor. In the much denser fluid beneath the crust, the short-range repulsive part of the neutron–neutron interaction does not favour this; instead, pairs have one quantum unit of orbital angular momentum, as in laboratory superfluid helium-3. Protons beneath the crust also pair into a liquid superconducting state if the local magnetic field is not prohibitively intense. Deeper in the star, any mesons present will probably form Bose condensates, another kind of superfluid. Even quark soup will pair up into an exotic state known as colour superconductivity<sup>5</sup>. Thus, at almost every depth within a neutron star, interpenetrating superfluids exist that flow without viscous drag, and that could conceivably participate in glitches.

The most popular model for radio-pulsar glitches holds the inner-crust neutron superfluid responsible<sup>4,6</sup>. One property of a rotating superfluid is that all of its vorticity — all circulation tendency in the flow — is concentrated in quantum vortex lines: multitudes of microscopic nodes, or holes, in the superfluid that thread through the fluid parallel to the rotation axis. As the superfluid in a star spins down, these quantum vortices migrate outward towards the equator and ultimately annihilate near the stellar surface. But vortices in the deep crust can become ‘pinned’, or stuck, to the crust lattice nuclei. This keeps the inner-crust superfluid rotating faster than the rest of the star, until the vortices come loose in a catastrophic unpinning event, observed as a pulsar glitch.

So far, so good. But in their study, Archibald and colleagues report an ‘anti-glitch’. Instead of an abrupt spin-up, the star abruptly spun down. (Such phenomena have been seen before, but only at a much smaller level<sup>7</sup>.) Despite searches with sensitive radio and X-ray telescopes, no surrounding afterglow was detected, arguing against a sudden particle outflow that could have carried off the anti-glitch’s angular momentum. The most likely

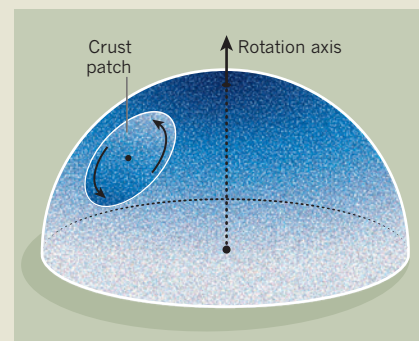
## BOX 1

## Idealized model for a magnetar anti-glitch

The star and its interior superfluid begin in co-rotation. The number of superfluid vortices penetrating the inner crust per unit area (blue shading) is highest at the north pole, but diminishes towards the equator owing to the curved stellar surface. A circular crust patch is then rotated (arrows) by plastic deformation, driven by stresses exerted by the twisted magnetic field below the crust. If this happens slowly and/or steadily, then its effect on the star’s spin-down history might not be obvious.

Assuming that the vortices remain pinned to nuclei in the crust, once the patch has turned through a significant angle ( $180^\circ$  in this illustration, much less in realistic applications<sup>9</sup>), the diminished number of vortices near the rotation axis will mean that the inner-crust superfluid is rotating more slowly than the rest of the star. A subsequent unpinning event would cause an anti-glitch.

Alternatively, if crust vortices are unevenly distributed as a result of previous deformations or non-uniform pinning, then a fast, magnetically driven twist, perhaps involving fracturing around a crust patch, would cause an anti-glitch whenever the net concentration of vortices near the rotation axis goes up. **R.C.D.**



inference seems to be that some superfluid component within the star was rotating more slowly than the crust before the anti-glitch and/or was torqued-up by a sudden internal rearrangement.

This is surprising, because the star in question, like all other solitary, magnetic neutron stars (including all radio pulsars), is spinning down monotonically, apart from occasional, ordinary glitches<sup>7</sup>. It is expected that internal superfluids can lag behind the general spin-down and act as faster-rotating ‘flywheels’. But how could a stellar superfluid come to rotate more slowly than the crust and/or get abruptly spun up?

The observed star is not a radio pulsar. It is a magnetar, an extremely magnetized neutron star with observable emissions powered by magnetic-field decay<sup>8</sup>. Magnetars are thought to be born spinning fast, with initial rotation periods of the order of several milliseconds. Their intense magnetism probably includes strongly wound-up interior-field components as a relic of this initial spin. The interior evolution of these objects is dominated by diffusing, changing magnetic fields. These exert stresses capable of moving material around, especially within radially concentric shells inside the stably stratified star — a possibility that may hold the key to the anti-glitch puzzle.

Box 1 describes an idealized model for how an anti-glitch could arise in the inner-crust superfluid<sup>9</sup>. A more promising alternative explanation involves the core. In moderately magnetized zones of the outer core, type II

proton superconductivity requires that all magnetic flux is concentrated in quantum flux tubes — microscopic nodes in the superconductor that are threaded by magnetic-field lines. These tubes resist passing through vortices in the interpenetrating neutron superfluid where vortices and tubes intersect<sup>10</sup>. Thus, many flux tubes together can exert compelling forces on vortices. Elsewhere in the core, or where type I proton superconductivity occurs, macroscopic flux structures with pronounced edges will also exert forces on vortices. As the star’s magnetism evolves, any gradual tendency of these flux systems to drive vortices away from the rotation axis will spin down the core neutron superfluid, potentially setting up conditions for an anti-glitch. A sudden ‘breakthrough’, or any instability by which vortices shift inward, would quickly spin up the superfluid and, as a result of angular-momentum conservation, spin down the rest of the star.

Perhaps of relevance is the fact that recent studies<sup>11,12</sup> have raised doubts about whether the inner-crust neutron superfluid is massive enough to account for ordinary radio-pulsar glitches. Therefore, some astrophysicists are already considering sub-crust superfluids as reservoirs for glitch angular momentum. However this issue is resolved, the anti-glitch will probably provide insight into the interiors of neutron stars and help to illuminate the strange life histories of magnetars. ■

Robert C. Duncan is in the Department of



Astronomy, University of Texas at Austin,  
Austin, Texas 78712, USA.  
e-mail: duncan@astro.utexas.edu

1. Archibald, R. F. *et al.* *Nature* **497**, 591–593 (2013).
2. Espinoza, C. M., Lyne, A. G., Stappers, B. W. & Kramer, M. *Mon. Not. R. Astron. Soc.* **414**, 1679–1704 (2011).
3. Yu, M. *et al.* *Mon. Not. R. Astron. Soc.* **429**, 688–724 (2013).
4. Pines, D. & Alpar, M. A. *Nature* **316**, 27–32 (1985).
5. Alford, M. G., Schmitt, A., Rajagopal, K. & Schafer, T. *Rev. Mod. Phys.* **80**, 1455–1515 (2008).

6. Anderson, P. W. & Itoh, N. *Nature* **256**, 25–27 (1975).
7. Iqdem, B., Baykal, A. & Inam, S. C. *Mon. Not. R. Astron. Soc.* **419**, 3109–3114 (2012).
8. Duncan, R. C. & Thompson, C. *Astrophys. J.* **392**, L9–L13 (1992).
9. Thompson, C. *et al.* *Astrophys. J.* **543**, 340–350 (2000).
10. Ruderman, M., Zhu, T. & Chen, K. *Astrophys. J.* **492**, 267–280 (1998).
11. Andersson, N., Glampedakis, K., Ho, W. C. G. & Espinoza, C. M. *Phys. Rev. Lett.* **109**, 241103 (2012).
12. Chamel, N. *Phys. Rev. Lett.* **110**, 011101 (2013).

E2 enzyme, it nevertheless promotes polySUMOylation. This may seem paradoxical, but Klug and colleagues show that enhanced polySUMOylation reflects the ability of Ubc9\*SUMO to act as a scaffold (Fig. 1).

An earlier study has shown<sup>6</sup> that Ubc9 can interact non-covalently with SUMO at a surface that is both spatially and functionally distinct from its catalytic site. Notably, this rear surface must be intact for Ubc9 to mediate polySUMOylation efficiently. Klug and co-authors show that, in this Ubc9\*SUMO interaction, the SUMO moiety binds to the rear surface of an active Ubc9 (which the authors call the enzyme's backside), which is linked to another SUMO through a thioester bond. At the same time, the catalytic site of Ubc9\*SUMO binds the consensus motifs of the thioester-linked SUMO, to place it in an optimal configuration for isopeptide-bond formation with the next SUMO protein as the latter is recruited by the catalytically active Ubc9.

This intricate mechanism shares some features with mechanisms reported earlier, in which E2 and E2-like proteins function not as enzymes but as scaffolds during the conjugation of target proteins. For example, one type of ubiquitin chain is assembled by an E2 enzyme in association with an inactive E2-like accessory protein<sup>7</sup>. The accessory protein binds an incoming ubiquitin so that the appropriate lysine residue of this ubiquitin is correctly oriented with respect to the catalytic site of the active E2 enzyme, ensuring that this lysine is

## BIOCHEMISTRY

# Rear view of an enzyme

**The enzyme Ubc9 mediates attachment of the small modifier protein SUMO to target proteins. It emerges that for optimal functioning — and for proper meiotic cell division — Ubc9 itself must be modified by SUMO.**

MARY DASSO

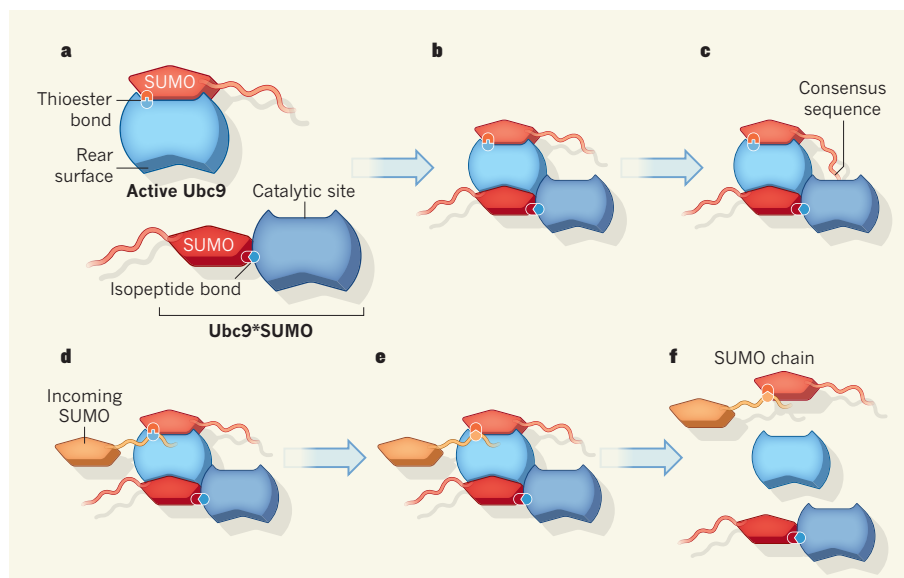
Chemical modification can change a protein's fate and behaviour. One such modification is SUMOylation, which involves covalent attachment of the small protein SUMO to target proteins. SUMOylated proteins can carry individual SUMO molecules or chains, with each modification resulting in a different fate. Typically, the addition of SUMO chains through the process of polySUMOylation marks the target for further tagging with chains of the SUMO-related protein ubiquitin and so for degradation<sup>1</sup>. Writing in *Molecular Cell*, Klug *et al.*<sup>2</sup> show that enhanced polySUMOylation is essential for meiotic cell division. The authors also describe how the balance is tipped between the addition of a single SUMO and a SUMO chain\*.

The attachment of SUMO to other proteins involves the sequential action of E1 and E2 enzymes<sup>1</sup>. The E1 enzyme Aos1/Uba2 binds to and activates SUMO, before establishing a thioester linkage between SUMO and the E2 enzyme Ubc9. This latter enzyme catalyses the formation of a stable isopeptide bond between the carboxy terminus of SUMO and lysine amino-acid residues in the target protein. This step frequently involves a third enzyme, E3 ligase.

Amino acids adjacent to the target protein's acceptor lysine are often important for Ubc9 activity, particularly when E3 enzymes are not involved. In fact, many target proteins carry a 'consensus' sequence that is preferentially used for SUMOylation in this situation. There are several such consensus sequences in the flexible amino-terminal domain of the SUMO protein found in the budding yeast *Saccharomyces cerevisiae*, and the lysines in them act

as the main sites for building SUMO chains<sup>3</sup>.

Klug *et al.* find that, remarkably, the activity of *S. cerevisiae* Ubc9 is controlled by its own SUMOylation. Previous work<sup>4,5</sup> had shown that yeast Ubc9 can be SUMOylated at two lysine residues near its own carboxy terminus, and that these modifications negatively regulate the ability of Ubc9 to conjugate SUMO to target proteins. The present paper demonstrates that although such SUMOylated Ubc9 (Ubc9\*SUMO) is catalytically inactive as an



**Figure 1 | Forging a link in the SUMO chain.** a, Attachment of the carboxy terminus of a SUMO protein to the catalytic site of Ubc9 through a thioester bond potentiates enzymatic activity of Ubc9. By contrast, SUMO conjugation close to the C terminus of Ubc9 through an isopeptide bond (Ubc9\*SUMO) blocks its enzymatic activity. Klug *et al.*<sup>2</sup> find that Ubc9\*SUMO promotes chain formation by acting as a scaffold. The SUMO moiety of Ubc9\*SUMO binds to the rear of the active Ubc9 (b). The catalytic site of Ubc9\*SUMO then recognizes a consensus sequence at the amino terminus of the SUMO attached to the active Ubc9, placing this SUMO in an optimal configuration for transfer to a target protein (c). The active Ubc9 recognizes an incoming SUMO (d) and catalyses formation of an isopeptide bond between the thioester-linked SUMO and a lysine with the N terminus of the incoming SUMO (e). Chain release follows (f).

\*This News & Views article was published online on 22 May 2013.

used to form the next link of the chain.

Additionally, mammalian Ubc9 forms an extremely stable complex with RanBP2, a nuclear-pore protein<sup>8</sup>. This complex acts as a composite E3 enzyme: the RanBP2-bound Ubc9 has a structural role, with a second Ubc9 catalysing isopeptide-bond formation between SUMO and its target protein. Intriguingly, the rear surface of Ubc9 that is stably bound to RanBP2 does not seem to be exposed, and so the orientation of the two Ubc9 molecules associated with RanBP2 probably differs from the arrangement of paired Ubc9 molecules described by Klug and colleagues. Collectively, these mechanisms suggest that E2 enzymes can be adapted for a variety of structural roles in conjugation pathways, and they hint that many more variations on this theme will be found.

Klug and co-authors also investigated the significance of Ubc9 SUMOylation during meiosis — the process by which a parental cell containing two sets of chromosomes eventually divides into four daughter cells, each carrying a single chromosome set. Yeast cells containing a Ubc9 mutant that could not be SUMOylated failed to form a synaptonemal complex, a meiotic structure that promotes both genetic exchange between chromosome pairs and correct chromosome segregation to daughter cells. In these mutants, individual

components of the synaptonemal complex showed altered concentrations as well as faulty incorporation into the complex. These findings build on earlier results<sup>9</sup> to indicate that the formation of SUMO chains is necessary for the synaptonemal complex to function, although many details of the meiotic defect remain unclear.

The same Ubc9 mutants grew normally under all the non-meiotic conditions that Klug *et al.* tested, and showed no increased sensitivity to stress conditions. This may imply either that normal growth and stress responses do not require polySUMOylation levels as high as those needed for meiosis, or that other mechanisms for achieving polySUMOylation can be invoked and are sufficient outside meiosis. The researchers note that wild-type Ubc9 becomes increasingly SUMOylated during meiosis, and it will be interesting to discover how this 'autoSUMOylation' is triggered to promote the polySUMOylation of downstream targets.

Ubc9 is the sole E2 enzyme in the SUMO pathway and so can act as a key regulatory point under circumstances in which the overall pattern of SUMOylation fluctuates in response to varying cellular conditions<sup>10,11</sup>. Although different consequences of autoSUMOylation on different domains of mammalian Ubc9 have been described<sup>5</sup>, it would be useful to

know whether Klug and colleagues' mechanism also applies in human cells. More generally, the authors' work suggests that there is much to learn about the feedback regulation of the SUMO pathway through SUMOylation and the many circumstances in which these mechanisms contribute to cellular function. ■

**Mary Dasso** is in the Program in Cellular Regulation and Metabolism, National Institute of Child Health and Human Development, Bethesda, Maryland 20892, USA.  
e-mail: dassom@mail.nih.gov

1. Gareau, J. R. & Lima, C. D. *Nature Rev. Mol. Cell Biol.* **11**, 861–871 (2010).
2. Klug, H. *et al. Mol. Cell* <http://dx.doi.org/10.1016/j.molcel.2013.03.027> (2013).
3. Bylebyl, G. R., Belichenko, I. & Johnson, E. S. *J. Biol. Chem.* **278**, 44113–44120 (2003).
4. Ho, C.-W., Chen, H.-T. & Hwang, J. *J. Biol. Chem.* **286**, 21826–21834 (2011).
5. Knipscheer, P. *et al. Mol. Cell* **31**, 371–382 (2008).
6. Knipscheer, P., van Dijk, W. J., Olsen, J. V., Mann, M. & Sixma, T. K. *EMBO J.* **26**, 2797–2807 (2007).
7. Eddins, M. J., Carlile, C. M., Gomez, K. M., Pickart, C. M. & Wolberger, C. *Nature Struct. Mol. Biol.* **13**, 915–920 (2006).
8. Werner, A., Flotho, A. & Melchior, F. *Mol. Cell* **46**, 287–298 (2012).
9. Watts, F. Z. & Hoffmann, E. *BioEssays* **33**, 529–537 (2011).
10. Hsieh, Y.-L. *et al. EMBO J.* **32**, 791–804 (2013).
11. Kelley, J. B. *et al. Mol. Cell. Biol.* **31**, 3378–3395 (2011).

## CANCER

# Drug for an 'undruggable' protein

**Scientists have long aimed to develop drugs against the cancer-associated protein KRAS, but without success. An approach that targets the oncoprotein's cellular localization reignites lost enthusiasm. SEE LETTER P.638**

NICOLE M. BAKER & CHANNING J. DER

**H**uman RAS genes have two claims to notoriety. First, they make up the most frequently mutated oncogene family in human cancer, having a prevalence of one in every three cases<sup>1</sup>. Second, despite more than three decades of intensive effort, no effective pharmacological inhibitor of the RAS oncoprotein has reached the clinic. So it is exciting that, on page 638 of this issue, Zimmermann *et al.*<sup>2</sup> report\* the identification and characterization of a small-molecule inhibitor that interferes with the localization of KRAS — the RAS isoform most commonly mutated in human cancers — to the plasma membrane surrounding cells<sup>3</sup>.

Following their synthesis in the cytoplasm,

RAS proteins are initially inactive<sup>4</sup>. They then undergo a series of rapid post-translational modifications that ensure their association with the inner leaflet of the plasma membrane, where these proteins exert their normal, as well as their cancer-associated, signalling activity. Therefore, most efforts aimed at anti-RAS drug discovery have involved indirect approaches to block the activities of proteins that either promote plasma-membrane association of RAS or are components of its downstream signalling pathway.

The key post-translational modification of RAS involves the addition of a 15-carbon farnesyl lipid tail in a reaction catalysed by the farnesyltransferase enzyme. This modification facilitates RAS association with membranes and is essential for proper RAS localization and activity, having prompted intensive efforts in the 1990s to develop

farnesyltransferase inhibitors (FTIs).

Despite promising results in preclinical studies, however, the results of clinical trials with FTIs were disappointing. The inhibitors blocked membrane association of the HRAS isoform, but lacked antitumour activity in cancers involving mutated KRAS (and NRAS). KRAS could still associate with the plasma membrane through an unexpected compensatory activity of the farnesyltransferase-related enzyme geranylgeranyltransferase-I, which modifies RAS with a geranylgeranyl, rather than a farnesyl, group. This discouraging outcome greatly dampened interest in targeting RAS — and, in particular, its membrane association — for cancer treatment. Instead, ongoing efforts have mainly focused on inhibitors of the RAF–MEK–ERK and the PI3K–AKT signalling cascades downstream of RAS.

Zimmermann *et al.* describe an approach aimed at disrupting KRAS membrane association that warrants reassessment of the current strategies. The authors identify and characterize a small-molecule inhibitor of PDEδ, a protein that can bind to and regulate the trafficking of RAS and RAS-related proteins to membrane compartments<sup>5–8</sup> (Box 1). Specifically, PDEδ contains a deep, hydrophobic pocket capable of binding the lipid moiety of farnesylated proteins, in particular RAS.

An earlier study<sup>8</sup> found that suppression of PDEδ levels disrupts RAS association with the plasma membrane and impairs the growth of RAS-mutant cancer cells. This finding

\*This article and the paper under discussion<sup>2</sup> were published online on 22 May 2013.



## BOX 1

## Inhibitors of the RAS signalling pathway

After their synthesis, inactive RAS proteins undergo a rapid series of post-translational modifications at a CaaX tetrapeptide motif on their carboxy terminus (C denotes the amino acid cysteine; aa, two saturated hydrocarbon (aliphatic) residues; and X, a variable residue). The enzymes that mediate this modification include farnesyltransferase (FTase), which attaches a 15-carbon farnesyl lipid to RAS. Other modifications to this lipid and to the CaaX motif follow, before the PDE $\delta$  protein is attached to the farnesylated tail of RAS, thereby facilitating RAS accumulation at the plasma membrane.

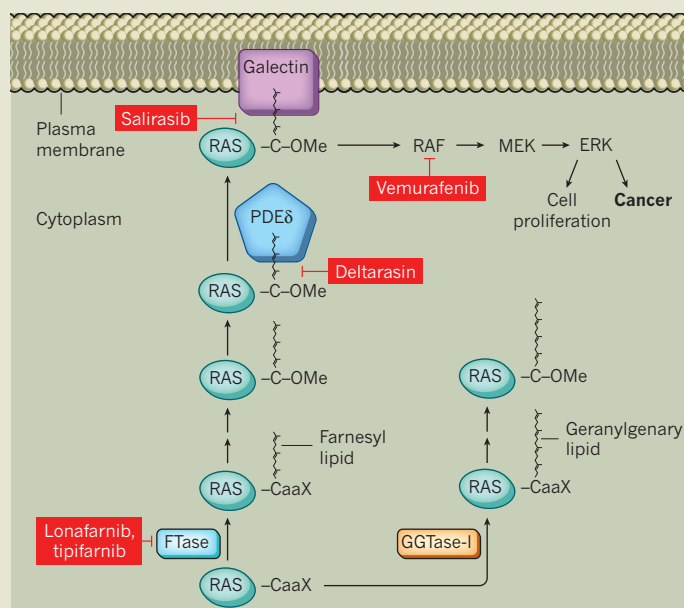
Pharmacological strategies to prevent activated RAS from reaching the plasma membrane include farnesyltransferase inhibitors such as lonafarnib and tipifarnib. Both of these are considered to be ineffective against KRAS

and NRAS, however, because they induce alternative modification of RAS proteins by geranylgeranyltransferase-I (GGTase-I). Lonafarnib is under clinical evaluation for

treating the genetic disorder progeria, and is showing highly promising results. Another agent, salirasib, is a farnesyl mimic that competes with active RAS for binding to

the oncoprotein's docking protein, galectin, on the plasma membrane, leading to degradation of active cytoplasmic RAS. Salirasib has shown promising results in a clinical trial for pancreatic cancer<sup>10</sup>. Zimmermann *et al.*<sup>2</sup> describe a new drug candidate, deltarasin, which inhibits PDE $\delta$  binding to farnesylated RAS.

In addition to these inhibitors of Ras association with the plasma membrane, more than 20 inhibitors that target the RAF–MEK–ERK signalling cascade downstream of RAS are under clinical evaluation. Of these, vemurafenib, a RAF inhibitor, has been approved for use in patients with BRAF-mutant metastatic melanoma, a skin cancer. **N.M.B. & C.J.D.**



prompted Zimmermann *et al.* to perform a high-throughput screen to identify small molecules that could block PDE $\delta$  association with the farnesylated tail of KRAS. After identifying several hits, they took a structure-based drug-design approach to develop their most promising compound, designated deltarasin.

The researchers' fluorescence microscopy experiments validate deltarasin's ability to block PDE $\delta$ –KRAS interaction in live cells. Following addition of 5-micromolar deltarasin to human KRAS-mutant pancreatic-cancer cell lines, PDE $\delta$  could no longer redistribute KRAS to the plasma membrane. Deltarasin also impaired the proliferative capacity of the pancreatic-cancer cell lines, which depends on signalling of mutant KRAS. Furthermore, it greatly reduced KRAS-dependent signalling events, such as phosphorylation of the ERK1 and ERK2 proteins. When the authors assessed the effects of deltarasin *in vivo*, in a mouse model of pancreatic ductal adenocarcinoma, they observed a dose-dependent reduction in tumour growth.

Recently, there has been a resurgence of interest in targeting RAS. This is due in part to the findings of cancer-genomics studies, which have reaffirmed KRAS mutations as the predominant oncogenic abnormalities in several cancers, including pancreatic, lung and colorectal cancers. However, the task of blocking mutant KRAS function with anticancer drugs

remains daunting, requiring identification of fresh strategies. Zimmermann and colleagues' finding, that pharmacologically disrupting the ability of PDE $\delta$  to promote plasma-membrane association of KRAS impairs the growth of KRAS-mutant pancreatic tumour cells, points to a provocative and innovative approach that may succeed where FTIs failed (Box 1).

Indeed, the efficacy of a PDE $\delta$  inhibitor will not be circumvented by alternative prenylation mechanisms, such as geranylgeranylation, that prevent FTIs from blocking KRAS (and NRAS) association with the plasma membrane. Nonetheless, as is the case with FTIs, inhibition of RAS regulation by PDE $\delta$  will probably have unforeseen consequences. PDE $\delta$  can interact with other farnesylated proteins, including farnesylated RAS-family proteins that act as tumour suppressors (for example, NOY2, also called DiRAS3). Inhibition of such beneficial proteins may lead to toxic effects in normal cells.

There are also conflicting observations regarding PDE $\delta$  selectivity for farnesylated and geranylgeranylated proteins<sup>6</sup>. If PDE $\delta$  is also required for the trafficking of geranylgeranylated proteins (such as Rho proteins), then additional off-target effects may be seen.

Another unresolved issue is exactly how dependent RAS proteins are on PDE $\delta$  for proper localization, because even in the absence of PDE $\delta$ , KRAS can bind to cell

membranes<sup>8</sup>. That KRAS deficiency is lethal, whereas PDE $\delta$  deficiency is not<sup>9</sup>, underscores the fact that at least some KRAS functions are PDE $\delta$  independent. So, although Zimmermann and co-authors' data are exciting, much remains to be learnt about PDE $\delta$  function and mutant-RAS dependency on it for proper subcellular localization. ■

**Nicole M. Baker and Channing J. Der** are in the Lineberger Comprehensive Cancer Center, Department of Pharmacology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7295, USA. e-mail: cjder@med.unc.edu

1. Cox, A. D. & Der, C. J. *Small GTPases* **1**, 2–27 (2010).
2. Zimmermann, G. *et al.* *Nature* **497**, 638–642 (2013).
3. Berndt, N., Hamilton, A. D. & Sebti, S. M. *Nature Rev. Cancer* **11**, 775–791 (2011).
4. Ahearn, I. M., Haigis, K., Bar-Sagi, D. & Philips, M. R. *Nature Rev. Mol. Cell Biol.* **13**, 39–51 (2012).
5. Philips, M. R. *Nature Cell Biol.* **14**, 128–129 (2012).
6. Nancy, V., Callebaut, I., El Marjou, A. & de Gunzburg, J. *J. Biol. Chem.* **277**, 15076–15084 (2002).
7. Ismail, S. A. *et al.* *Nature Chem. Biol.* **7**, 942–949 (2011).
8. Chandra, A. *et al.* *Nature Cell Biol.* **14**, 148–158 (2012).
9. Zhang, H. *et al.* *Proc. Natl Acad. Sci. USA* **104**, 8857–8862 (2007).
10. Laheru, D. *et al.* *Invest. New Drugs* **30**, 2391–2399 (2012).

C.J.D. declares competing financial interests. See [go.nature.com/s5fsuu](http://go.nature.com/s5fsuu) for details.

# The Norway spruce genome sequence and conifer genome evolution

Lists of authors and their affiliations appear at the end of the paper

Conifers have dominated forests for more than 200 million years and are of huge ecological and economic importance. Here we present the draft assembly of the 20-gigabase genome of Norway spruce (*Picea abies*), the first available for any gymnosperm. The number of well-supported genes (28,354) is similar to the >100 times smaller genome of *Arabidopsis thaliana*, and there is no evidence of a recent whole-genome duplication in the gymnosperm lineage. Instead, the large genome size seems to result from the slow and steady accumulation of a diverse set of long-terminal repeat transposable elements, possibly owing to the lack of an efficient elimination mechanism. Comparative sequencing of *Pinus sylvestris*, *Abies sibirica*, *Juniperus communis*, *Taxus baccata* and *Gnetum gnemon* reveals that the transposable element diversity is shared among extant conifers. Expression of 24-nucleotide small RNAs, previously implicated in transposable element silencing, is tissue-specific and much lower than in other plants. We further identify numerous long (>10,000 base pairs) introns, gene-like fragments, uncharacterized long non-coding RNAs and short RNAs. This opens up new genomic avenues for conifer forestry and breeding.

Gymnosperms are a group of land plants comprising the extant taxa, cycads, *Ginkgo*, gnetophytes and conifers. Gymnosperms first appeared more than 300 million years ago (Myr ago)<sup>1</sup>, well before the angiosperm lineage separated from the stem group of extant gymnosperms<sup>2</sup>. The major radiation of conifer families occurred 250–65 Myr ago<sup>3</sup>, and during their evolution the morphology of conifers has changed relatively little. There are approximately 630 conifer species, representing about 70 currently recognized genera, which dominate many terrestrial ecosystems, especially in the Northern Hemisphere. Conifers also dominated both before and after the major mass extinction events at the end of the Permian and Cretaceous periods, around 250 and 65 Myr ago, respectively. Conifers are of immense ecological and economic value; coniferous forests cover enormous areas in the Northern Hemisphere, and conifers are keystone species in many other ecosystems. Conifers contribute a large fraction of terrestrial photosynthesis and biomass, and the cultural and economic values of conifers are also paramount; early civilizations used conifers for firewood, tools and artefacts and today several national economies depend on commodities produced from conifers. However, despite their abundance and importance, our understanding of conifer genomes is limited. Most conifers have 12 ( $2n = 24$ ) chromosomes, probably reflecting the ancestral karyotype<sup>4</sup>, which are typically of similar size, each being roughly comparable to the size of the human genome, and containing high proportions of repetitive elements<sup>5,6</sup>. The gene space of conifer genomes has not been well characterized, although several reports have suggested that gene families in conifers may be larger than their angiosperm counterparts<sup>7</sup> and that conifer genomes contain numerous pseudogenes<sup>8</sup>.

Because their genomes are among the largest—typically 20–30 gigabases pairs (Gb)—of all organisms, genome-wide analyses of conifers are particularly challenging. Thus, no full genome sequence of a gymnosperm species is available at present, whereas 30 angiosperm and more basal plant genomes have been sequenced. However, size is not the only challenge to sequencing presented by conifer genomes. Conifers are typically outbreeding, produce wind-dispersed pollen, have very large effective population sizes, and their genomes are highly heterozygous, although their nucleotide substitution rates are lower than those of most angiosperms<sup>8,9</sup>, perhaps owing to long life-span (decades to centuries). Furthermore, inbreeding depression

negates the production of inbred lines that could facilitate genome assembly.

The availability of conifer genome sequences would enable comparative analyses of genome architecture and the evolution of key traits for seed plants, including flower or fruit development and life history (perennial versus annual), and help to determine how and why conifer genomes became so large. To address these issues and aid forest tree breeding, biodiversity and conservation studies by, for example, enabling the genome-wide design of genetic markers, we used data from massively parallel DNA sequencing to assemble a draft of the 20-Gb nuclear genome of Norway spruce (*Picea abies* (L.) Karst), one of the most widespread, ecologically and economically important plants in Europe. We analysed the protein-coding and non-coding fractions of the genome and compared them to the low-coverage draft genome assemblies of five other gymnosperms—Scots pine (*P. sylvestris*), Siberian fir (*A. sibirica*), juniper (*J. communis*), yew (*Taxus baccata*) and *Gnetum gnemon*—to gain insight into conifer genome evolution.

## Sequencing and assembly

We sequenced a 43-year-old root-grafted copy of the *P. abies* clone Z4006, which originated from a tree in Ragunda, central Sweden, collected in 1959. Many copies of this clone are available in clone archives and seed orchards, and it has been extensively used in Swedish breeding programs. We estimated its genome size to be 19.6 Gb ( $C = 20.02 \pm 0.95$  pg (mean  $\pm$  s.d.); Supplementary Information 1.1), in accordance with previous reports<sup>10</sup>.

*De novo* sequencing and assembly of large, repeat-containing, heterozygous genomes remains challenging. To assemble the *P. abies* genome, we developed a hierarchical strategy combining fosmid pools<sup>11</sup> with both haploid and diploid whole genome shotgun (WGS) data, and RNA sequencing (RNA-Seq) data<sup>12–14</sup> (Supplementary Information 1.2–1.3). The resulting assembly (*P.abies* 1.0) included 4.3 Gb in >10-kilobase (kb) scaffolds (Table 1), and we estimated that approximately 63% of protein-coding genes<sup>15</sup> were fully covered (>90% of their length), and 96% partially covered (>30% of their length) within single scaffolds (Supplementary Information 1.4). By mapping diploid reads to the *P.abies* 1.0 assembly, the single nucleotide



polymorphism (SNP) frequency was estimated to be 0.77% and the short insertion/deletion (indel) frequency to be 0.05% (Supplementary Information 1.5).

The chloroplast genome (124 kb) revealed considerable structural variation within the genus *Picea* (Supplementary Information 1.6). The draft mitochondrial genome (>4 Mb) was among the largest reported for plants and was rich in short open-reading frames (ORFs), which appeared to be gene remnants derived from repeat-driven mitochondrial rearrangements<sup>16</sup> (Supplementary Information 1.7).

## Presence of long introns and gene-like fragments

We generated >1 billion RNA-Seq reads and used transcript assemblies of these in combination with public expressed sequence tags (ESTs) and transcripts to perform *ab initio* prediction of protein-coding genes, which identified a high confidence set of 28,354 loci with >70% coverage by supporting evidence from the total set of 70,968 predicted loci. A notable characteristic of the predicted gene structures was the presence of numerous long introns (Fig. 1b), with mean intron length being higher than in most available plant genomes, although similar to the repeat-rich genomes of *Vitis vinifera* and *Zea mays*<sup>17,18</sup>. The longest intron in the high-confidence genes was 68 kb (Supplementary Table 2.6), and 2,384 high-confidence genes contained 2,880 longer than 5-kb introns (20 of which we confirmed by PCR amplification; Supplementary Information 2.14), 2,679 of which contained a repeat, suggesting that repeat insertions account for intron expansion. By contrast, exon size was consistent among the species considered (Supplementary Information 2.6.3). Numerous genes (~30%) remained split across scaffolds owing to assembly fragmentation, and as such, the longest introns were not represented in the *P. abies* 1.0 assembly. Long introns (either individual or cumulative intron length) did not influence expression levels (Fig. 1c) and introns containing repeats have not contracted despite a lack of recent repeat activity (see below).

Analysis of gene families in the high-confidence gene set and seven sequenced plant genomes (five angiosperms: *Arabidopsis thaliana*, *Populus trichocarpa*, *Vitis vinifera*, *Oryza sativa* and *Zea mays*, and two basal plants: *Selaginella moellendorffii* and *Physcomitrella patens*) identified 1,021 *P. abies*-specific gene families (Fig. 1a and Supplementary Information 2.8). *P. abies*-specific families included over-representation of Gene Ontology categories involved in DNA repair and methylation of DNA and chromatin (Supplementary Information 2.8). As for most draft genomes, these results probably

overestimate gene numbers<sup>19</sup> and will be refined as we further improve the genome assembly.

A common mechanism leading to genome size expansion is the occurrence of a whole genome duplication (WGD) event. We calculated the number of synonymous substitutions per synonymous site ( $K_s$ ) of paralogues within the high-confidence genes but found no evidence for any recent WGD; there was a clear, exponential decay in the number of retained paralogues with increasing  $K_s$  values (Supplementary Information 2.9 and Supplementary Fig. 2.6). However, a population dynamics model that takes into account both small- and large-scale modes of gene duplication<sup>20</sup> suggested the presence of a small peak (around  $K_s$  of 1.1), which, considering the slow substitution rate of conifers, might represent the ancient WGD predating the divergence of angiosperms and gymnosperms (350 Myr ago<sup>21</sup>).

Previous examinations of small genomic subsets indicated that conifer genomes contain numerous pseudogenes<sup>5,6,22,23</sup>. The gene-like fraction of the *P. abies* 1.0 assembly was identified by alignment of RNA-Seq reads and *de novo* assembled transcripts (Supplementary Information 2.10). Within this subset of the genome, loci with valid spliced alignments of *de novo* assembled transcripts or the presence of a high-confidence gene were also identified. The high-confidence gene set represented 27 Mb of protein-coding sequence, whereas 72 Mb of regions were identified with a valid spliced alignment or a high-confidence gene. In stark contrast, 524 Mb of gene-like regions were identified by less stringent alignments. The presence of such a large gene-like fraction lacking predicted gene structures supports the presence of numerous pseudogenes.

Recent ENCODE publications<sup>24,25</sup> characterized numerous long non-coding RNA (lncRNA) loci in the human genome, but this class of RNA remains largely uncharacterized in plants. Using short-read *de novo* transcript assemblies, 13,031 spruce-specific and 9,686 conserved intergenic lncRNAs were identified (Supplementary Information 2.4.3). In common with the ENCODE results, *P. abies* lncRNA loci contained fewer exons, were shorter (Fig. 1c), and had more tissue-specific expression than protein-coding loci (Supplementary Fig. 2.8).

There has been conflicting evidence about the presence of 24-nucleotide short RNAs (sRNAs) in gymnosperms<sup>26–29</sup>, a class of sRNA that silence transposable elements by the establishment of DNA methylation<sup>30</sup>. Across 22 samples, we identified numerous 24-nucleotide sRNAs, but these were highly specific to reproductive tissues, largely associated with repeats but present at substantially lower levels than in angiosperms (Fig. 1d and Supplementary Fig. 2.10). By contrast, 21-nucleotide sRNAs were associated with genes, repeats and promoters/untranslated regions (UTRs) (Fig. 1d). *De novo* microRNA (miRNA) prediction identified 2,719 loci, including 20 known miRNA families, with target sites predicted within the high-confidence gene set for 1,378 of these (Supplementary Information 2.13). Furthermore, 55 known miRNA families had >5 aligned sRNA reads and mature miRNAs, representing 49 known families aligned to the genome (Supplementary Information 2.13).

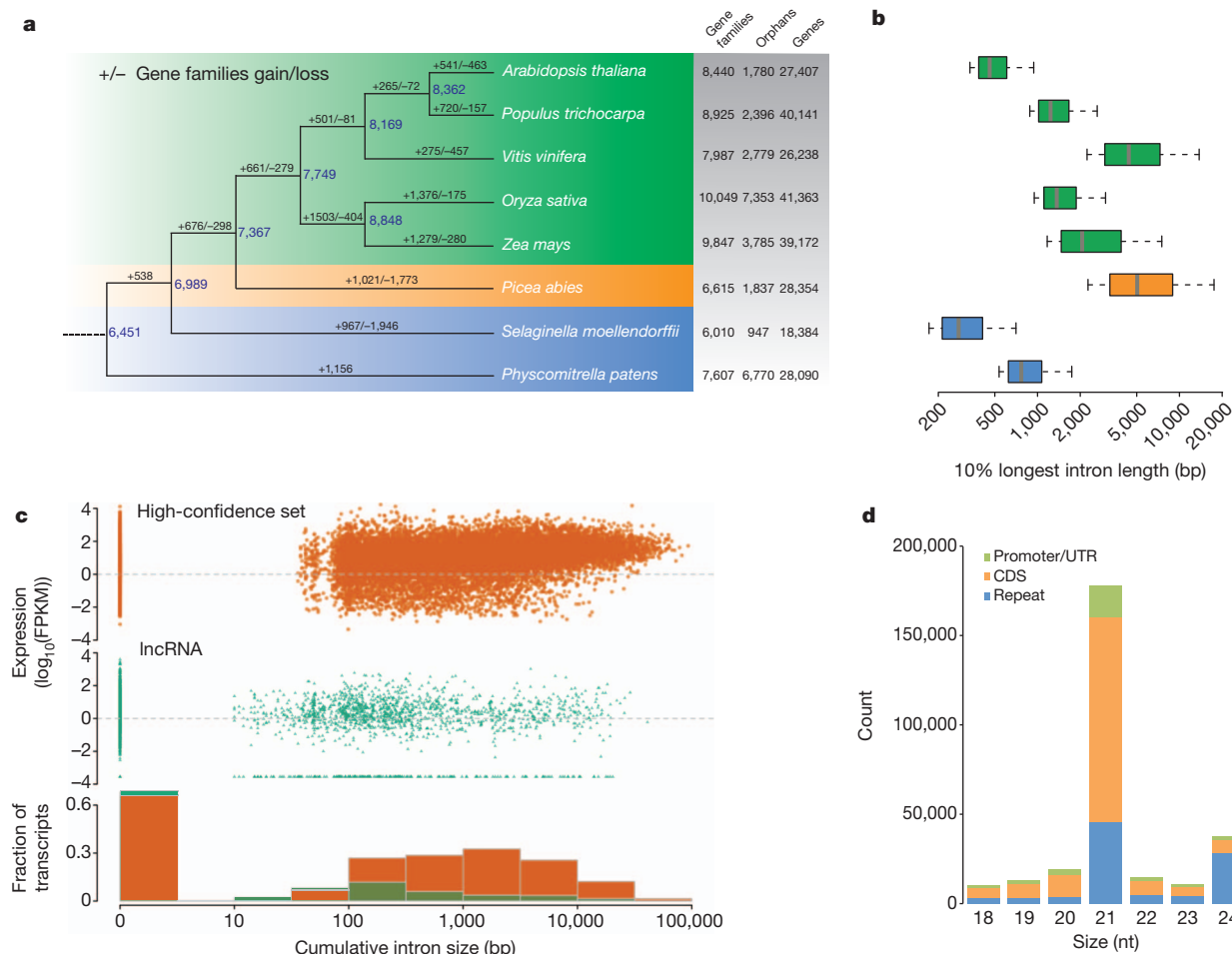
## Conifer genomes grew by insertion of LTR-RT elements

We constructed a manually curated library of 1,773 repetitive sequences, approximately half of which could be assigned to known transposable element repeat families (Supplementary Information 3.1–3.3). Long terminal repeat-retrotransposons (LTR-RTs) comprised the most abundant fraction of transposable elements, with the *Ty3/Gypsy* superfamily being more abundant than the *Ty1/Copia* superfamily (Fig. 2a and Table 1). We also identified and characterized transposable elements using 454 reads from randomly sheared genomic DNA in five other gymnosperms (*P. sylvestris*, *A. sibirica*, *J. communis*, *T. baccata* and *G. gnemon*) and, in all six species, LTR-RTs were the most abundant class (Fig. 2a, Supplementary Information 4.1 and Supplementary Table 3.1).

**Table 1 | Characteristics of the *P. abies* genome**

Genome	
Size (1n)	19.6 Gb
Karyotype	2n = 24
GC content	37.9%
High-copy repeat content*	
LTR_Gypsy/Copia/unclassified	35%/16%/7%
LINE	1%
DNA transposable element	1%
Unclassified	10%
Genes and gene-like fragments†	2.4%
Assembly	
Size in scaffolds >200 bp/>10 kb	12 Gb/4.3 Gb
N50/NG50	4,869 bp/721 bp
Annotation	
High confidence gene set	28,354
Genes with >5-kb introns	8.4%
Avg. exon/intron size	312 bp/1,017 bp
Avg. gene density	1 gene in 705 kb
Transposable element genes	284,587
Non-coding loci	
lncRNA (unique/conserved)	13,031/9,686
miRNA ( <i>de novo</i> predicted)	2,719

\*Inferred from unassembled reads. †Including pseudogenes, excluding transposable elements.



**Figure 1 | The gene-space and transcribed fraction of the *P.abies* 1.0 assembly.** **a**, Gene family loss and gain in eight sequenced plant genomes (*Arabidopsis thaliana*, *Populus trichocarpa*, *Vitis vinifera*, *Oryza sativa*, *Zea mays*, *Picea abies*, *Selaginella moellendorffii* and *Physcomitrella patens*). Gene families were identified using TribeMCL (inflation value 4), and the DOLLOP program from the PHYLIP package was used to determine the minimum gene set for ancestral nodes of the phylogenetic tree. We used plant genome annotations filtered to remove transposable elements. 'Orphans' refers to gene families containing only a single gene. Blue numbers indicate the number of

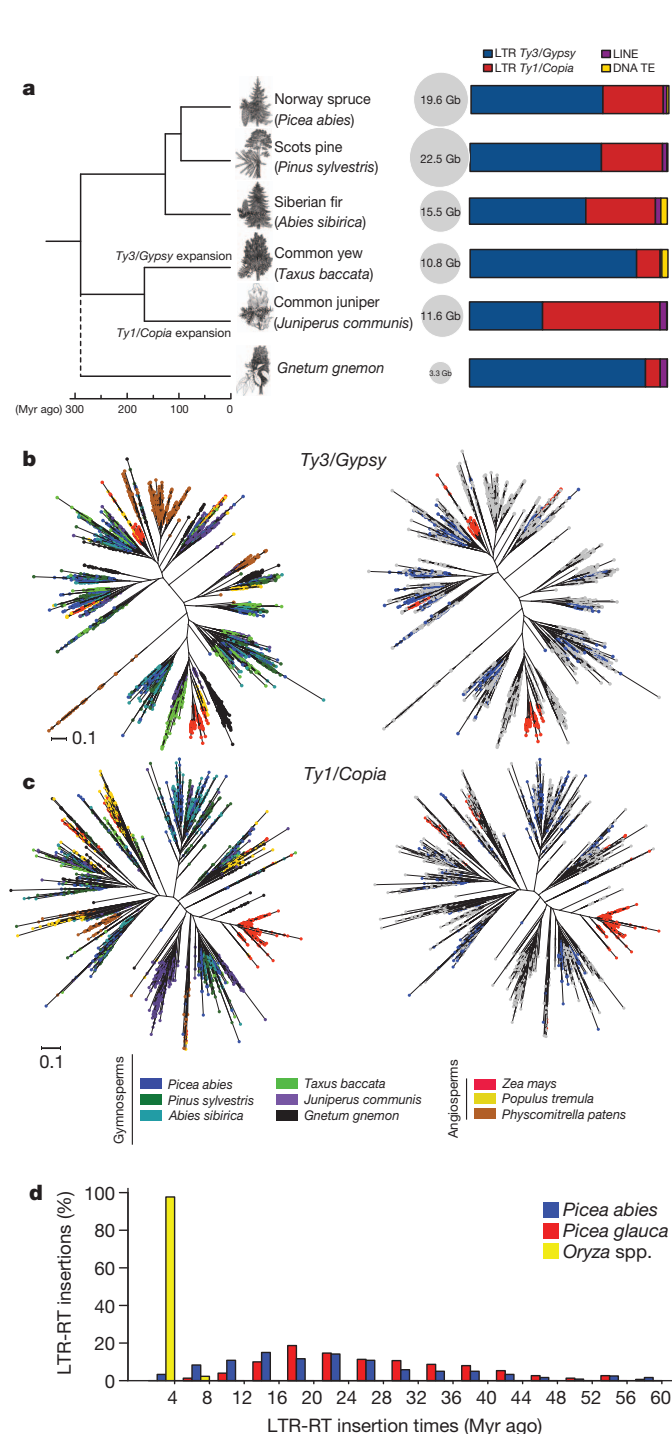
gene families. **b**, Boxplot representation of length distribution for the 10% longest introns in the same eight genomes. **c**, Scatter plots of cumulative intron length against  $\log_{10}$  expression calculated as fragments per kilobase per million mapped reads (FPKM) for high-confidence gene loci (top, coloured orange) and green for IncRNA loci (middle, shaded green). The bottom panel shows a histogram of cumulative intron size in the two sets of loci. **d**, Distribution of small (18–24-nucleotide (nt)) RNAs and their co-alignment-based colocation to genomic features (repeats, high-confidence genes and their promoter/UTRs). CDS, coding sequence.

To trace the history of transposable elements in vascular plants we inferred phylogenies of a domain of the reverse transcriptase genes of both *Ty1/Copia* and *Ty3/Gypsy* elements. The phylogenies revealed several diverse and ancient transposable element subfamilies, present in almost all of the examined conifer genera, whereas only a few subfamilies were expanded in the angiosperm genomes (Fig. 2b, c and Supplementary Information 3.11). Most internal clades with significant bootstrap support were consistently species-specific, indicating that most expansions of extant transposable element families occurred after divergence. Two species-specific amplification bursts were evident: a *Ty1/Copia* family in *J. communis* and a *Ty3/Gypsy* family in *T. baccata*. We used complete LTR-RTs from *P. abies* and *P. glauca* to investigate further the timing of conifer transposable element insertions<sup>31</sup> (Supplementary Information 3.4–3.8). In contrast to a similar set of elements identified in *Oryza sativa* and *O. glaberrima* (Fig. 2d), we detected no evidence of recent activity (that is, less than 5 Myr ago) in *P. abies*. Instead, insertions seem to have occurred over several tens of millions of years (older insertions are more likely to escape detection). Analysis of 68 orthologous transposable element insertions in *P. abies* and *P. glauca* further supported this: 63 insertions apparently predated divergence, and only five occurred after the lineages separated 13–20 Myr ago (Supplementary Information 3.9).

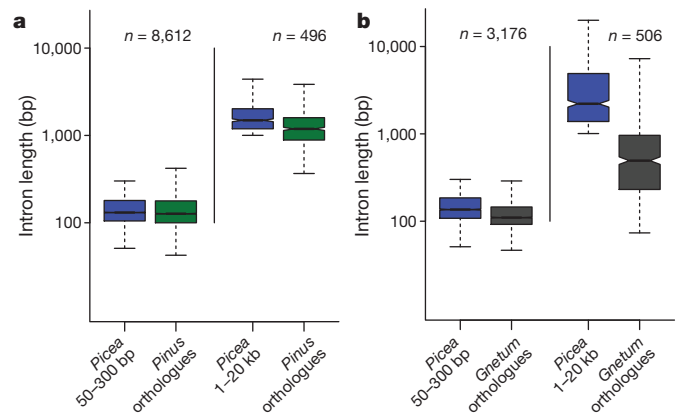
We clustered LTRs of complete elements to identify transposable element families<sup>32</sup>. More than 86% of the elements remained as singletons, indicating that LTR-RTs are quite divergent and that there are several low-abundance families. We searched three LTR-RT families for signatures of unequal intra-element recombination events in scaffolds >50 kb and 20 complete fosmids<sup>33</sup>. For families ALISEI, 3K05 and 4D08\_5 we identified 21, 22 and 39 complete elements, and four, five and no solo LTRs, respectively (Supplementary Information 3.10). Although this data set is limited, the analysis suggested that LTR-RT-related sequences might be removed less frequently by unequal recombination than in other plant genomes. The ratio of solo-LTRs to complete elements in *P. abies* is ~1:9, whereas in *A. thaliana*, rice and barley it is 1:1 (ref. 33), 0.6:1 (ref. 34) and 16:1 (for the abundant BARE 1 element<sup>35</sup>), respectively. Taken together, these findings indicated that the extant set of transposable elements in *P. abies* accumulated slowly over tens or hundreds of millions of years, mainly by the insertion of LTR-RT elements with limited transposable element removal.

An analysis of introns across taxa provided further insight into the genome of the last common ancestor to the conifers. We identified orthologues of normal sized (50–300 bp) and long (1–20 kb) introns in





**Figure 2 | Conifer genomes contain expansions of a diverse set of LTR-RTs.** **a**, Distribution of different classes of transposable elements from six gymnosperm species. The figure is based on the total fraction of transposable elements (TE) identified and grouped into different classes from the different species. Genome sizes of the six species are given in circles and their phylogenetic relationship is shown, with tentative dating of divergence times (x-axis) based on 64 chloroplast genes over 39 species and five fossil calibration points. **b**, **c**, Heuristic neighbour-joining trees constructed from 5,922 sequences similar to the Ty3/Gypsy (**b**) and 3,052 sequences similar to the Ty1/Copia (**c**) reverse transcriptase domain from nine plant species. The trees to the right have only sequences from *P. abies* and *Z. mays* coloured, whereas the grey dots are the uncoloured versions of the other species represented on the left. **d**, Distributions of insertion times calculated for LTR-RTs in *Picea abies*, *Picea glauca* and *Oryza glaberrima*/O. *sativa*, using mutation rates (per base per year) of  $2.2 \times 10^{-9}$  for the *Picea* spp. and  $1.8 \times 10^{-8}$  for O. *glaberrima*<sup>50</sup>.



**Figure 3 | Intron sizes are conserved among gymnosperms.** **a**, **b**, Intron size comparisons between *P. abies*, *P. sylvestris* (**a**) and *G. gnemon* (**b**), respectively. Orthologues of introns that were categorised as short (50–300 bp) or long (1–20 kb) in *P. abies* were identified in *P. sylvestris* and *G. gnemon*, and the corresponding intron size was scored.

spruce within draft genome assemblies of *P. sylvestris* and *G. gnemon* (Supplementary Information 4.2). Introns identified as orthologous to a long intron in *P. abies* were also atypically long (Fig. 3a, b), suggesting that intron expansions started early in the history of conifers.

### The evolution of important conifer traits

Two major differences between angiosperms and gymnosperms are their contrasting reproductive development and the development of water-conducting xylem cells. We therefore manually identified *P. abies* loci homologous to genes known to be centrally involved in these processes in angiosperms.

In angiosperms, homologues of the *A. thaliana* phosphatidylethanolamine-binding protein (PEBP) FLOWERING LOCUS T (FT) are key activators of flowering. It has been suggested that gymnosperms lack orthologues of FT genes, instead containing a group of FT/TFL1-like genes that probably act as flowering repressors<sup>36,37</sup>. We identified four putative FT/TFL1-like genes in the *P. abies* 1.0 genome that have not been previously described and confirmed that the genome does indeed lack FT-like genes (Supplementary Information 5.1).

MADS-box genes are involved in controlling most aspects of angiosperm development<sup>38</sup>. A total of 278 sequences with clear homology to MADS boxes were identified in the *P. abies* 1.0 assembly (Supplementary Information 5.2), 41 of which had transcript support. Type I and II MADS-box genes are distinguished in plants. Only 5% of the identified MADS boxes were of type I (Supplementary Fig. 5.2.), the lowest percentage of potential type I genes recorded in any plant genome. Type II MADS-box genes are subdivided into about a dozen ancient clades. We observed remarkable expansions in the TM3-like (or SOC1-like), STMADS11-like and TM8-like gene clades in *P. abies*. Because members of these gene clades are involved in vegetative development and phase changes such as the floral transition in angiosperms<sup>39</sup>, we propose that the expansion of these gene clades has contributed to the evolution of developmental phase changes in gymnosperms.

The xylem tissue of most gymnosperms comprises a single water-transporting cell type, tracheids. By contrast, the xylem tissue of angiosperm species contains fibres, originating from tracheids that have to a large extent lost the capacity to conduct mass water flows, and vessels that have taken over the water-transport function in the stem. Formation of vessels is controlled by the VASCULAR NAC DOMAIN (VND) gene family, which has seven members in *A. thaliana*<sup>40</sup>. We detected two VND genes in *P. abies* (Supplementary Information 5.3), suggesting that co-option and expansion of the VND gene family in vessel formation might have been important for angiosperm evolution.

## A model for conifer genome evolution

We propose the following model of conifer genome evolution. After the lineage that led to angiosperms had branched off and the most recent common ancestor of extant conifers had been established, the 12 ancestral chromosomes expanded at a slow and steady rate due to the activity of a diverse set of *Gypsy* and *Copia* LTR transposable elements that are largely shared among extant conifers. The expansion started early and, in contrast to angiosperms genomes in which this has been counteracted by efficient recombination mechanisms<sup>33</sup> resulting in only smaller transposable element subsets remaining following recent bursts of activity<sup>41,42</sup>, these elements have remained in the genome. We propose that mechanisms for transposable element removal (for example by unequal recombination) have been less active in conifers than in most other organisms<sup>43</sup>, and our data suggest that the insertion of transposable elements into genes gave rise to large introns, and (combined with other mechanisms) abundant pseudogenes. Each chromosome has grown to a similar size—perhaps limited by physical constraints on, for instance, chromosomal replication—with genes separated by large regions of transposable-element-rich, highly polymorphic non-protein-coding regions with low recombination frequencies. The gradual increase in size, the lack of WGDs and a predominately out-crossing mating system have probably also buffered conifer genomes against chromosomal rearrangements (WGD reduces sensitivity to aneuploidy), thereby maintaining synteny over large phylogenetic distances<sup>44</sup>.

Some angiosperms, such as cereals, also have large genomes but it seems as if the “one way ticket towards genome obesity”<sup>45</sup> that is barely recognizable in angiosperms prevails in conifers. The underlying mechanism remains unclear, but the low frequency of 24-nucleotide sRNAs, their role in methylation of repeats and their restriction to reproductive tissues may have influenced the process. However, considering the effect of methylation patterns on recombination rates<sup>46</sup> and the fact that 24-nucleotide sRNAs trigger methylation, such low recombination frequencies would more likely result from hypermethylation<sup>47</sup>. A state of ‘genome paralysis’ could potentially have been triggered once an obesity threshold was reached. In the angiosperm lineage, the occurrence of a number of WGDs probably increased diversification potential, allowing morphological innovation (for example, the origin of flowers and fruits) and facilitating speciation<sup>46,48,49</sup>. By contrast, the conserved genome structure resulting from the paucity of genome rearrangements and lack of WGDs in conifers probably limited the evolution of reproductive barriers (resulting in relatively low rates of speciation), and may explain the high degrees of conservation through time and low morphological diversity. Nevertheless, these processes do not seem to affect fitness as conifers dominate many ecosystems, probably because they contain high degrees of standing genetic variation, allowing them to occupy very wide ecological niches in climatic regions where other plant species are less competitive. The future availability of additional gymnosperm genome sequences will allow further exploration of the unique processes that have driven their evolution and facilitate improvement of this important species.

## METHODS SUMMARY

We shotgun-sequenced 450 fosmid pools containing around 100–6,000 fosmids per pool (Supplementary Table 1.4). Each fosmid pool was assembled and scaffolded individually. Fosmid pool scaffolds larger than 1 kb (~6.7 Gb in total) were merged<sup>12</sup> (Supplementary Information 1.3) with a 38× haploid WGS assembly (~9.8 Gb in total, derived from ~600 ng of DNA extracted from a single megagametophyte). We subsequently performed scaffolding<sup>13</sup> using WGS libraries of five different insert sizes (0.3, 0.65, 2.4, 4.4 and 10.4 kb) from diploid tissue. We further increased assembly contiguity of protein-coding regions by scaffolding using a set of ~38 million unassembled (after digital normalization) RNA-Seq read-pairs generated from 22 samples (Supplementary Information 1.3).

*Ab initio* prediction of protein-coding genes was performed using ESTs from numerous conifer species, our own short-read *de novo* transcript assemblies and

proteins from other plant species as supporting evidence (Supplementary Information 2.6). Predicted loci were used to perform gene family analysis and to examine the  $K_s$  substitution rates of paralogues to identify evidence for a recent WGD event. *De novo* transcript assemblies were used to identify lncRNA, and sRNA sequencing was performed and used for *de novo* miRNA prediction.

Repeated sequences were identified *de novo* using 454 reads longer than 700 bp generated from randomly sheared genomic DNA. Candidates were characterized using similarity searches at the nucleotide and amino acid level against public and custom collections of plant transposable element sequences. Complete LTR-RTs were identified using a combination of *de novo* searches and manual inspection.

WGS assemblies from shallow sequencing (3.8–12.5×) of *P. sylvestris*, *A. sibirica*, *J. communis*, *T. baccata* and *G. gnemon* were produced using the CLC Bio *de novo* assembler.

For website and accession number information, see Supplementary Information 6.

Received 10 February; accepted 22 April 2013.

Published online 22 May 2013.

- Stewart, W. N. & Rothwell, G. W. *Paleobotany and the Evolution of Plants* (Cambridge Univ. Press, 1993).
- Savard, L. et al. Chloroplast and nuclear gene sequences indicate late Pennsylvanian time for the last common ancestor of extant seed plants. *Proc. Natl Acad. Sci. USA* **91**, 5163–5167 (1994).
- Leslie, A. B. et al. Hemisphere-scale differences in conifer evolutionary dynamics. *Proc. Natl Acad. Sci. USA* **109**, 16217–16221 (2012).
- Flory, W. S. Chromosome numbers and phylogeny in the gymnosperms. *J. Arnold Arb.* **17**, 82–87 (1936).
- Morse, A. M. et al. Evolution of genome size and complexity in *Pinus*. *PLoS ONE* **4**, e4332 (2009).
- Kovach, A. et al. The *Pinus taeda* genome is characterized by diverse and highly diverged repetitive sequences. *BMC Genomics* **11**, 420 (2010).
- Ahuja, M. R. & Neale, D. B. Evolution of genome size in conifers. *Silvae Genet.* **54**, 126–137 (2005).
- Buschiazio, E., Ritland, C. E., Bohlmann, J. & Ritland, K. Slow but not low: genomic comparisons reveal slower evolutionary rate and higher dN/dS in conifers compared to angiosperms. *BMC Evol. Biol.* **12**, 8 (2012).
- Jaramillo-Correa, J. P., Verdu, M. & González-Martínez, S. C. The contribution of recombination to heterozygosity differs among plant evolutionary lineages and life-forms. *BMC Evol. Biol.* **10**, 22 (2010).
- Murray, B. G. Nuclear DNA amounts in gymnosperms. *Ann. Bot. (Lond.)* **82**, 3–15 (1998).
- Zhang, G. et al. The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* **490**, 49–54 (2012).
- Vicedomini, R., Vezzi, F., Scalabrini, S., Arvestad, L. & Policriti, A. GAM-NGS: genomic assemblies merger for next generation sequencing. *BMC Bioinformatics* **14**, S6 (2013).
- Sahlén, K., Street, N., Lundberg, J. & Arvestad, L. Improved gap size estimation for scaffolding algorithms. *Bioinformatics* **28**, 2215–2222 (2012).
- Vezzi, F., Narzisi, G. & Mishra, B. Feature-by-feature-evaluating de novo sequence assembly. *PLoS ONE* **7**, e31002 (2012).
- Ralph, S. G. et al. A conifer genomics resource of 200,000 spruce (*Picea* spp.) ESTs and 6,464 high-quality, sequence-finished full-length cDNAs for Sitka spruce (*Picea sitchensis*). *BMC Genomics* **9**, 484 (2008).
- Mackenzie, S. A. in *Plant Mitochondria* (ed. Logan, D. C.) 36–49 (Blackwell, 2007).
- Messing, J. et al. Sequence composition and genome organization of maize. *Proc. Natl Acad. Sci. USA* **101**, 14349–14354 (2004).
- Jaillon, O. et al. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* **449**, 463–467 (2007).
- Bennetzen, J. L., Coleman, C., Liu, R., Ma, J. & Ramakrishna, W. Consistent over-estimation of gene number in complex plant genomes. *Curr. Opin. Plant Biol.* **7**, 732–736 (2004).
- Vanneste, K., Van de Peer, Y. & Maere, S. Inference of genome duplications from age distributions revisited. *Mol. Biol. Evol.* **30**, 177–190 (2013).
- Jiao, Y. et al. Ancestral polyploidy in seed plants and angiosperms. *Nature* **473**, 97–100 (2011).
- García-Gil, M. R. Evolutionary aspects of functional and pseudogene members of the phytochrome gene family in Scots pine. *J. Mol. Evol.* **67**, 222–232 (2008).
- Magbanua, Z. V. et al. Adventures in the enormous: a 1.8 million clone BAC library for the 21.7 Gb genome of loblolly pine. *PLoS ONE* **6**, e16214 (2011).
- Bánfai, B. et al. Long noncoding RNAs are rarely translated in two human cell lines. *Genome Res.* **22**, 1646–1657 (2012).
- Derrien, T. et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* **22**, 1775–1789 (2012).
- Dolgosheina, E. V. et al. Conifers have a unique small RNA silencing signature. *RNA* **14**, 1508–1515 (2008).
- Morin, R. D. et al. Comparative analysis of the small RNA transcriptomes of *Pinus contorta* and *Oryza sativa*. *Genome Res.* **18**, 571–584 (2008).
- Wan, L.-C. et al. Identification and characterization of small non-coding RNAs from Chinese fir by high throughput sequencing. *BMC Plant Biol.* **12**, 146 (2012).
- Zhang, J. et al. Dynamic expression of small RNA populations in larch (*Larix leptolepis*). *Planta* **237**, 89–101 (2013).



30. Henderson, I. R. & Jacobsen, S. E. Epigenetic inheritance in plants. *Nature* **447**, 418–424 (2007).
31. Sanmiguel, P. & Bennetzen, J. L. Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. *Ann. Bot. (Lond.)* **82**, 37–44 (1998).
32. Wicker, T. *et al.* A unified classification system for eukaryotic transposable elements. *Nature Rev. Genet.* **8**, 973–982 (2007).
33. Devos, K. M., Brown, J. K. M. & Bennetzen, J. L. Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. *Genome Res.* **12**, 1075–1079 (2002).
34. Vitte, C. & Panaud, O. Formation of solo-LTRs through unequal homologous recombination counterbalances amplifications of LTR retrotransposons in rice *Oryza sativa* L. *Mol. Biol. Evol.* **20**, 528–540 (2003).
35. Vicent, C. M. *et al.* Retrotransposon BARE-1 and its role in genome evolution in the genus *Hordeum*. *Plant Cell* **11**, 1769–1784 (1999).
36. Karlgren, A. *et al.* Evolution of the PEBP gene family in plants: functional diversification in seed plant evolution. *Plant Physiol.* **156**, 1967–1977 (2011).
37. Klintonäs, M., Pin, P. A., Benlloch, R., Ingvarsson, P. K. & Nilsson, O. Analysis of conifer *FLOWERING LOCUS T/TERMINAL FLOWER1*-like genes provides evidence for dramatic biochemical evolution in the angiosperm *FT* lineage. *New Phytol.* **196**, 1260–1273 (2012).
38. Gramzow, L. & Theissen, G. A hitchhiker's guide to the MADS world of plants. *Genome Biol.* **11**, 214 (2010).
39. Smaczniak, C. *et al.* Developmental and evolutionary diversity of plant MADS-domain factors: insights from recent studies. *Development* **139**, 3081–3098 (2012).
40. Kubo, M. *et al.* Transcription switches for protoxylem and metaxylem vessel formation. *Genes Dev.* **19**, 1855–1860 (2005).
41. Piegu, B. *et al.* Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. *Genome Res.* **16**, 1262–1269 (2006).
42. Hawkins, J. S., Kim, H., Nason, J. D., Wing, R. A. & Wendel, J. F. Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium*. *Genome Res.* **16**, 1252–1261 (2006).
43. Bennetzen, J. L., Ma, J. & Devos, K. M. Mechanisms of recent genome size variation in flowering plants. *Ann. Bot. (Lond.)* **95**, 127–132 (2005).
44. Pavy, N. *et al.* A spruce gene map infers ancient plant genome reshuffling and subsequent slow evolution in the gymnosperm lineage leading to extant conifers. *BMC Biol.* **10**, 84 (2012).
45. Bennetzen, J. L. & Kellogg, E. A. Do plants have a one way ticket to genomic obesity? *Plant Cell* **9**, 1509–1514 (1997).
46. Van de Peer, Y., Fawcett, J. A., Proost, S., Sterck, L. & Vandepoele, K. The flowering world: a tale of duplications. *Trends Plant Sci.* **14**, 680–688 (2009).
47. Fedoroff, N. V. Presidential address. Transposable elements, epigenetics, and genome evolution. *Science* **338**, 758–767 (2012).
48. Van de Peer, Y. A mystery unveiled. *Genome Biol.* **12**, 113 (2011).
49. Soltis, D. E. *et al.* Polyploidy and angiosperm diversification. *Am. J. Bot.* **96**, 336–348 (2009).
50. Ma, J. & Bennetzen, J. L. Rapid recent growth and divergence of rice nuclear genomes. *Proc. Natl Acad. Sci. USA* **101**, 12404–12410 (2004).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** The authors would like to acknowledge support from the Knut and Alice Wallenberg Foundation. Additional funding was provided in particular by the Swedish Research Council (VR), the Swedish Governmental Agency for Innovation Systems (Vinnova), the Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning (Formas), the Swedish foundation for Strategic Research (SSF), the Government of Canada through Genome Canada, by Genome British Columbia, and by Genome Quebec, Science for Life Laboratory and the National Genomics Infrastructure (NGI), Sweden. We also acknowledge Skogforsk for the *P. abies* genetic material, UPPMAX for computational infrastructure, CLC Bio for assembly software development and Lucigen for fosmid-pool method development.

**Author Contributions** B.N. and N.R.S. are joint first authors, and A.W., A.Z., Y.-C.L. and D.G.S. are joint second authors, who contributed to most parts of the work. F.V., A.A., N.D., R.V., K.S. and E.S. contributed to the assembly and sequence analysis; S.G. to repeat analysis; N.D., M.E., L.G., M.Ku., T.N., Å.O., G.T., H.T., P.Z. and B.Z. to quality control of the assembly and analysis of gene families; K.H., J.H., O.K., M.Kä. and T.S. to the sequencing; L.K. to analysis of the mitochondrial genome; M.Ko. and N.R. to generation of fosmid pools; J.Lut., F.L., C.T.-L. and K.V. to analysis of the sequences of *P. abies* and other conifers; C.R. and J.S. to production of BAC sequences; Z.-Q.W. to analysis of the chloroplast genome and J.A.R. determined the genome size. L.A., R.B., J.Boh., J.Bou.,

R.G.G., T.R.H., P.d.J., J.M., M.M., K.R., B.S., S.L.T., Y.V.d.P. and B.A. contributed to the design and supervision of various parts of the research. O.N. headed and P.K.I. managed the project, J.Lun. coordinated the sequencing and assembly activities, and S.J. the bioinformatics activities. B.N., N.R.S., A.Z., O.N., P.K.I., J.Lun. and S.J. wrote and edited most of the manuscript. All authors commented on the manuscript.

**Author Information** Raw data and assemblies are available from the ConGenIE (Conifer Genome Integrative Explorer) web resource (<http://congenie.org>), as well as the European Bioinformatics Institute (EMBL) and European Nucleotide Archive (ENA); see Supplementary Information 6.1 for accession numbers. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to O.N. (ove.nilsson@slu.se), P.K.I. (par.ingvarsson@emg.umu.se), J.Lun. (joakim.lundeberg@scilifelab.se) or S.J. (stefan.jansson@umu.se).



This work is licensed under a Creative Commons Attribution-NonCommercial-Share Alike 3.0 Unported licence. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-sa/3.0>

Björn Nystedt<sup>1</sup>, Nathaniel R. Street<sup>2</sup>, Anna Wetterbom<sup>3</sup>, Andrea Zuccolo<sup>4,5</sup>, Yao-Cheng Lin<sup>6</sup>, Douglas G. Scofield<sup>2,7</sup>, Francesco Veczi<sup>8</sup>, Nicolas Delhomme<sup>2</sup>, Stefania Giacomello<sup>4,9</sup>, Andrey Alexeyenko<sup>10</sup>, Riccardo Vicedomini<sup>4,9</sup>, Kristoffer Sahlin<sup>8</sup>, Ellen Sherwood<sup>1</sup>, Malin Elfstrand<sup>11</sup>, Lydia Gramzow<sup>12</sup>, Kristina Holmberg<sup>10</sup>, Jimmie Hällman<sup>10</sup>, Olivier Keech<sup>2</sup>, Lisa Klasson<sup>13</sup>, Maxim Koriabine<sup>14</sup>, Melis Kucukoglu<sup>15</sup>, Max Käller<sup>10</sup>, Johannes Luthman<sup>3</sup>, Fredrik Lysholm<sup>3</sup>, Totte Niittylä<sup>15</sup>, Åke Olson<sup>11</sup>, Nemanja Rilakovic<sup>10</sup>, Carol Ritland<sup>16</sup>, Josep A. Rosselló<sup>17,18</sup>, Juliana Sena<sup>19</sup>, Thomas Svensson<sup>20</sup>, Carlos Talavera-López<sup>3</sup>, Günter Theißen<sup>12</sup>, Hannele Tuominen<sup>2</sup>, Kevin Vanneste<sup>6</sup>, Zhi-Qiang Wu<sup>7</sup>, Bo Zhang<sup>2</sup>, Philipp Zerbe<sup>21</sup>, Lars Arvestad<sup>8,22</sup>, Rishikesh Bhalerao<sup>15</sup>, Joerg Bohlmann<sup>16,21</sup>, Jean Bousquet<sup>19</sup>, Rosario Garcia Gil<sup>15</sup>, Torgeir R. Hvidsten<sup>2,23</sup>, Pieter de Jong<sup>14</sup>, John MacKay<sup>19</sup>, Michele Morgante<sup>4,9</sup>, Kermit Ritland<sup>16</sup>, Björn Sundberg<sup>15</sup>, Stacey Lee Thompson<sup>7</sup>, Yves Van de Peer<sup>6</sup>, Björn Andersson<sup>3</sup>, Ove Nilsson<sup>15</sup>, Pär K. Ingvarsson<sup>7</sup>, Joakim Lundeberg<sup>10</sup> & Stefan Jansson<sup>2</sup>

<sup>1</sup>Department of Biochemistry and Biophysics, Science for Life Laboratory, Stockholm University, Box 1031, 171 21 Solna, Sweden. <sup>2</sup>Umeå Plant Science Centre, Department of Plant Physiology, Umeå University, 901 87 Umeå, Sweden. <sup>3</sup>Department of Cell and Molecular Biology, Science for Life Laboratory, Karolinska Institutet, Box 1031, 171 77 Stockholm, Sweden. <sup>4</sup>Istituto di Genomica Applicata, Via J. Linussio 51, 33100 Udine, Italy. <sup>5</sup>Institute of Life Sciences, Scuola Superiore Sant'Anna, 56127 Pisa, Italy. <sup>6</sup>Department of Plant Systems Biology (VIB) and Department of Plant Biotechnology and Bioinformatics (Gent University), Technologiepark 927, 9052 Gent, Belgium. <sup>7</sup>Umeå Plant Science Centre, Department of Ecology and Environmental Science, Umeå University, 901 87 Umeå, Sweden. <sup>8</sup>School of Computer Science and Communication, Science for Life Laboratory, KTH Royal Institute of Technology, Box 1031, 171 21 Solna, Sweden. <sup>9</sup>Università degli Studi di Udine, Via delle Scienze 208, 33100 Udine, Italy. <sup>10</sup>School of Biotechnology, Science for Life Laboratory, KTH Royal Institute of Technology, Box 1031, 171 21 Solna, Sweden. <sup>11</sup>Department of Forest Mycology and Plant Pathology, Uppsala Biocentre, Swedish University of Agricultural Sciences, Box 7026, 750 07 Uppsala, Sweden. <sup>12</sup>Department of Genetics, Friedrich-Schiller-University Jena, Philosophenweg 12, 07743 Jena, Germany. <sup>13</sup>Molecular Evolution, Department of Cell and Molecular Biology, Uppsala University, Husargatan 3, 752 37 Uppsala, Sweden. <sup>14</sup>BACPAC Resources, Children's Hospital of Oakland Research Institute, Bruce Lyon Memorial Research Building, Oakland, California 94609, USA. <sup>15</sup>Umeå Plant Science Centre, Department of Forest Genetics and Plant Physiology, Swedish University of Agricultural Sciences, 901 83 Umeå, Sweden. <sup>16</sup>Department of Forest and Conservation Sciences, University of British Columbia, 2424 Main Mall, Vancouver, British Columbia V6T 1Z4, Canada. <sup>17</sup>Jardí Botànic, Universitat de València, c/Quart 80, 46008 Valencia, Spain. <sup>18</sup>Marimurtra Botanical Garden, Carl Faust Fdn, 17300 Blanes, Spain. <sup>19</sup>Canada Research Chair in Forest and Environmental Genomics, Centre for Forest Research and Institute for Systems and Integrative Biology, Université Laval, Québec, Québec G1V 0A6, Canada. <sup>20</sup>Department of Biosciences and Nutrition, Science for Life Laboratory, Karolinska Institutet, Box 1031, 171 21 Solna, Sweden. <sup>21</sup>Michael Smith Laboratories, University of British Columbia, 321-2185 East Mall, Vancouver, British Columbia V6T 1Z4, Canada. <sup>22</sup>Swedish e-Science Research Center, Department Numerical Analysis and Computer Science, Stockholm University, Box 1031, 171 21 Solna, Sweden. <sup>23</sup>Department of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, 1432 Ås, Norway.

# The importance of mixed selectivity in complex cognitive tasks

Mattia Rigotti<sup>1,2,3</sup>, Omri Barak<sup>1,†</sup>, Melissa R. Warden<sup>4,5</sup>, Xiao-Jing Wang<sup>2,6</sup>, Nathaniel D. Daw<sup>2,3</sup>, Earl K. Miller<sup>4</sup> & Stefano Fusi<sup>1</sup>

**Single-neuron activity in the prefrontal cortex (PFC) is tuned to mixtures of multiple task-related aspects. Such mixed selectivity is highly heterogeneous, seemingly disordered and therefore difficult to interpret. We analysed the neural activity recorded in monkeys during an object sequence memory task to identify a role of mixed selectivity in subserving the cognitive functions ascribed to the PFC. We show that mixed selectivity neurons encode distributed information about all task-relevant aspects. Each aspect can be decoded from the population of neurons even when single-cell selectivity to that aspect is eliminated. Moreover, mixed selectivity offers a significant computational advantage over specialized responses in terms of the repertoire of input-output functions implementable by readout neurons. This advantage originates from the highly diverse nonlinear selectivity to mixtures of task-relevant variables, a signature of high-dimensional neural representations. Crucially, this dimensionality is predictive of animal behaviour as it collapses in error trials. Our findings recommend a shift of focus for future studies from neurons that have easily interpretable response tuning to the widely observed, but rarely analysed, mixed selectivity neurons.**

Neurophysiology experiments in behaving animals are often analysed and modelled with a reverse engineering perspective, with the more or less explicit intention to identify highly specialized components with distinct functional roles in the behaviour under study. Physiologists often find the components they are looking for, contributing to the understanding of the functions and the underlying mechanisms of various brain areas, but they are also bewildered by numerous observations that are difficult to interpret. Many cells, especially in higher-order brain structures like the prefrontal cortex (PFC), often have complex and diverse response properties that are not organized anatomically, and that simultaneously reflect different parameters. These neurons are said to have mixed selectivity to multiple aspects of the task. For instance, in rule-based sensory-motor mapping tasks (such as the Wisconsin card sorting test), the response of a PFC cell may be correlated with parameters of the sensory stimuli, task rule, motor response or any combination of these<sup>1,2</sup>. The predominance of these mixed selectivity neurons seems to be a hallmark of PFC and other brain structures involved in cognition. Understanding such neural representations has been a major conceptual challenge in the field.

To characterize the statistics and understand the functional role of mixed selectivity, we analysed neural activity recorded in the PFC of monkeys during a sequence memory task<sup>3</sup>. Motivated by recent theoretical advances in understanding how machine learning principles play out in the functioning of neuronal circuits<sup>4–10</sup>, we devised a new analysis of the recorded population activity. This analysis revealed that the observed mixed selectivity can be naturally understood as a signature of the information-encoding strategy of state-of-the-art classifiers like support vector machines<sup>11</sup>. Specifically we found that (1) the populations of recorded neurons encode distributed information that is not contained in classical selectivity to individual task aspects, (2) the recorded neural representations are high-dimensional, and (3) the dimensionality of the recorded neural representations predicts behavioural performance.

## Dimensionality and mixed selectivity

The dimensionality of a neural representation in a given time bin is a property of a set of vectors, each of which represents the firing rates of  $N$  recorded neurons in one experimental condition. The pattern of activity encoded in each such vector can be thought of as a point in an  $N$ -dimensional space. Over a set of such points, the dimensionality we refer to is defined as the minimal number of coordinate axes that are needed to specify the position of all points (Supplementary Methods M.1). For example, if all points are on a line, then their dimensionality is one, as one appropriately aligned axis is sufficient to determine their position. The dimensionality of the neural representations recorded during an experiment is then the dimensionality generated by the patterns of activity observed in each of the different experimental conditions (for example, all combinations of sensory stimuli and behavioural responses).

High-dimensional neural representations have the desirable property of allowing simple readouts such as linear classifiers to implement a large set of input-output relations. Model circuits that rely on such high-dimensional representations can generate very rich dynamics and solve complex tasks<sup>5–10</sup>, and this same property is exploited in contemporary machine learning techniques such as support vector machines.

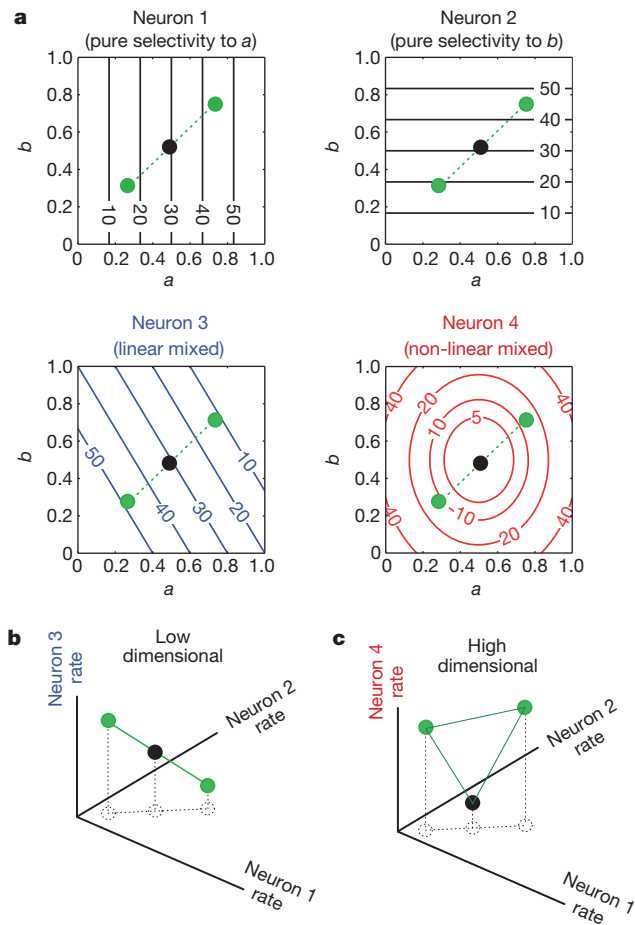
This dimensionality is related to the mixed selectivity of neuronal responses. This is because a set of neurons whose responses are selective only to individual task-relevant aspects, or even to linear sums of multiple aspects (linear mixed selectivity), can only generate low-dimensional representations. Higher dimensional representations can be produced by including neurons whose responses cannot be explained as a linear sum of aspect-related responses, that is, neurons with nonlinear mixed selectivity (Fig. 1a, b).

## Mixed selectivity in a memory task

Monkeys were trained to remember the identity and order of presentation of two objects sequentially displayed on a screen. Their

<sup>1</sup>Center for Theoretical Neuroscience, Columbia University College of Physicians and Surgeons, New York, New York 10032, USA. <sup>2</sup>Center for Neural Science, New York University, New York, New York 10003, USA. <sup>3</sup>Department of Psychology, New York University, New York, New York 10003, USA. <sup>4</sup>The Picower Institute for Learning and Memory & Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. <sup>5</sup>Department of Bioengineering, Stanford University, Stanford 94305, California, USA. <sup>6</sup>Department of Neurobiology, Kavli Institute for Neuroscience, Yale University School of Medicine, New Haven, Connecticut 06525, USA. <sup>†</sup>Present address: Department of Physiology, Technion Medical School, Haifa, 31096, Israel.

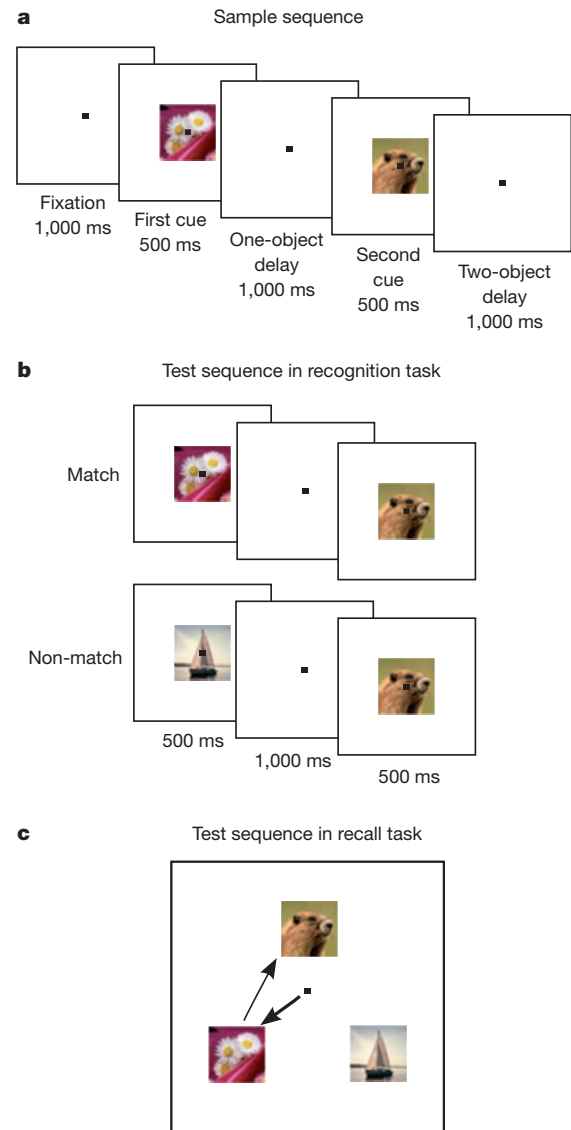




**Figure 1 | Low and high-dimensional neural representations, and mixed selectivity.** **a**, Contour plots of the responses (spikes per s) of four hypothetical neurons to two continuous parameters that characterize two task-relevant aspects (*a*, *b*, varying between 0 and 1) corresponding to relevant stimulus features (for example, contrast and orientation). Neurons 1, 2 are pure selectivity neurons, selective to individual parameters (*a* and *b*, respectively). Neuron 3 is a linear mixed selectivity neuron: its response is a linear combination of the responses to parameters *a* and *b*. Neuron 4 is a nonlinear mixed selectivity neuron: its response cannot be explained by a linear superposition of responses to the individual parameters. The green circles indicate the responses to three sensory stimuli parameterized by three *a*, *b* combinations. **b**, The responses of the pure and linear mixed selectivity neurons from **a** in the space of activity patterns (the axes indicate the firing rates of the neurons) elicited by the three stimuli indicated by the green circles in **a** lie on a line, therefore spanning a low-dimensional space. **c**, As in **b**, with the third neuron being the nonlinear mixed selectivity Neuron 4 in **a**. The representations of the stimuli lie on a plane, no longer being confined on a line. This higher dimensionality has an important role when the activity is read out by linear classifiers, because they can only separate the input space into classes that are separable by a plane (in general by a hyper-plane). This limits the implementable classifications (See Supplementary Section S.1). For example, in **b** it is impossible for any linear classifier to respond to the darker central circle and not to the other two. But it is possible in **c**, for instance for a linear classifier corresponding to an appropriately positioned horizontal plane.

memory was then tested after a delay (the two-object delay period) either through a recognition or through a recall task, which were interleaved in blocks of 100–150 trials (Fig. 2 and refs 3, 12 for more details).

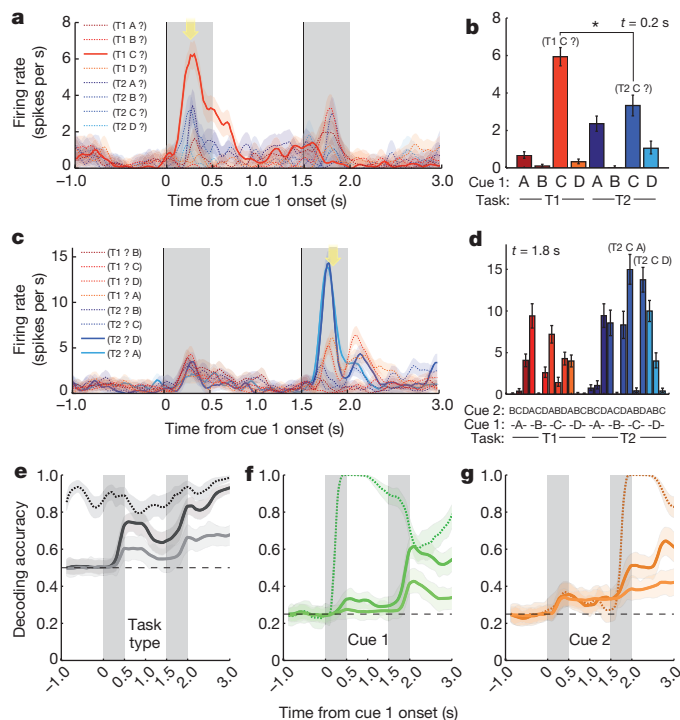
We analysed the activity of 237 lateral PFC neurons (area 46) recorded in two monkeys during the trial epochs that precede and include the two-object delay period. In these epochs, each trial was characterized by a condition defined by three relevant aspects: the identity of the two visual objects and the task type. The first cue was randomly selected from a predefined set of four objects. The second cue was randomly chosen among the three remaining objects.



**Figure 2 | Behavioural task from ref. 3.** **a**, Sample sequence: each trial began when the monkeys grasped a bar and achieved central fixation. A first sample object was followed by a brief delay (the one-object delay), then a second sample object (different from the first sample object), then another delay (the two-object delay). **b**, Recognition task: the sample sequence was followed by a test sequence, which was either a match to the sample sequence, in which case the monkeys were required to release the bar, or a non-match, in which case the monkeys were required to hold the bar until a matching sequence appeared. **c**, Recall task: the sample sequence was followed by an array of three objects that included the two sample objects. Monkeys were required to make a sequence of saccades in the correct order to the two sample objects. Recognition and recall task trials were interleaved in blocks of 100–150 trials.

Finally, the task type was either recognition or recall. More generally, for other experimental protocols, every situation (condition) would be characterized by the values of a set of discrete or continuous variables. We refer to them as to the task-relevant aspects.

The analysis of recorded single-neuron responses shows that the majority of neurons are selective to at least one of the three task-relevant aspects in one or more epochs<sup>3</sup>. A large proportion of neurons, moreover, show nonlinear mixed selectivity (Supplementary Section S.2). Figure 3a–d show two examples of nonlinear mixed selectivity neurons. Figure 3a shows a cell that is selective to a mixture of cue 1 identity and task type: it responds to object C when presented as a first cue, more strongly during the recognition task. The neuron of Fig. 3c, d is mostly selective to objects A and D when presented as



**Figure 3 | Mixed selectivity in recorded single-cell activity and population decoding.** **a**, Average firing rate of a sample neuron (Gaussian smoothing with 100-ms window, shaded area indicates s.e.m.). Colours denote different combinations of task type and sample cues (condition), indicated in parenthesis (task type, first cue, second cue). The '?' indicates that cue 2 identities were averaged over. This neuron preferentially responds to object C as first cue in task 1 blocks (recognition task). **b**, Peri-condition histogram (PCH): average firing rate in a 100-ms time bin ( $\pm$  s.e.m.) at the yellow arrow in **a** for different conditions. The response to object C as first cue is significantly different for the two task types ( $P < 0.05$ , two-sample  $t$ -test). **c**, **d**, Same as **a**, **b**, for a different neuron with preference for object A and D as second objects during task 2 trials (recall task). **e–g**, Comparison of population decoding accuracy for task type (**e**), cue 1 (**f**) and cue 2 (**g**) before (dashed) and after (solid) removing classical selectivity. Dashed lines: average trial-by-trial cross-validated decoding accuracy of the decoder reading out the firing rate of 237 neurons in different independent time bins. Curves represent the average decoding accuracy over 1,000 partitions of the data into training and test set (shaded areas show 95% confidence intervals). Horizontal dashed lines indicate chance level. Solid lines: decoding accuracy after the removal of classical selectivity for 237 (bright) and 1,000 resampled neurons (dark) (see Supplementary Methods M.6). **e**, Accuracy in decoding task type from neurons whose selectivity to task type was removed. The decoding accuracy is initially at chance level, but steadily grows above chance level as the complexity of the task and the number of conditions increases. **f**, **g**, Analogous plots for the decoding accuracy of cue 1,2 identity, when instead selectivity to cue 1,2 was removed.

second stimuli, but only during the recall task and when they are preceded by object C.

### Information encoded by mixed selectivity

A neuron is conventionally said to be selective to a task-relevant aspect if it responds differentially to the values of the parameters characterizing that aspect. A neuron is, for instance, selective to task type, if its average responses in recall and recognition task trials are significantly different. The operation of averaging over all conditions corresponding to a particular task type (for each task type there are 12 possible combinations of the sample visual objects) may, however, result in discarding important information. The responses in individual conditions could encode information about task type through the nonlinear interactions between the cue and the task type aspects, which manifest themselves as nonlinear mixed selectivity. This suggests that if nonlinear mixed selectivity is sufficiently diverse across

neurons, the information about task type could be extracted from the covariance between neuronal responses across different conditions, even when individual neurons are not 'classically' selective to task type (that is, the average responses to recall and recognition tasks are not significantly different). Information could in other words be distributed across the neural population, even when it is not present in individual cells (see also refs 13, 14 for recent discussions on distributed codes.)

To show that this is indeed the case, we manipulated the data to remove the classical selectivity to a given task aspect and then tested whether that task aspect could still be decoded from the population. Classical selectivity to a given task-relevant aspect is removed from every recorded neuron by adding noise that equalizes average responses, preserving the differences between the individual conditions (Supplementary Methods M.3). In Fig. 3e–g we compared the cross-validated trial-by-trial population-decoding accuracy before and after removing classical selectivity. Neurons that were not recorded simultaneously were combined as pseudo-simultaneous population activity patterns as explained in Supplementary Methods M.5. The temporal correlations that were neglected with this procedure do not seem to appreciably affect the decoding accuracy (Supplementary Section S.4).

Before removing classical selectivity, the maximal cross-validated decoding accuracy peaks close to 100% for all task-relevant aspects (Fig. 3e). Both the identity and the temporal order of the visual objects could also be decoded (see Supplementary Section S.6). Crucially, all task-relevant aspects can be decoded even when classical selectivity is removed. The first panel of Fig. 3e shows the accuracy of decoding task type from the intact population and after removing classical selectivity to task type from all neurons. Note that removing classical selectivity causes a larger drop of decoding accuracy in the early epochs of the trial. As the trial progresses and more visual cues are memorized (that is, the task becomes more complex), the accuracy progressively increases towards the values of the intact population. Moreover, the decoding accuracy increases as the number of neurons read out by the decoder increases. We estimated the decoding accuracy for larger neural populations by resampling the recorded neurons and randomly relabelling the identities of the visual objects, so as to obtain responses whose activity have the same statistics as the recorded ones. For example, a new neuron could be obtained by assigning the activity of a recorded neuron in response to objects A, B, C, D to the trials in which the objects were B, D, A, C (see Supplementary Methods M.3 and M.6 for more details). Similar results hold after removing the classical selectivity to cue 1 and cue 2 (Fig. 3f, g), or when we removed the classical selectivity by subtracting from the neural activity the linear mixed selectivity component (Supplementary Section S.3).

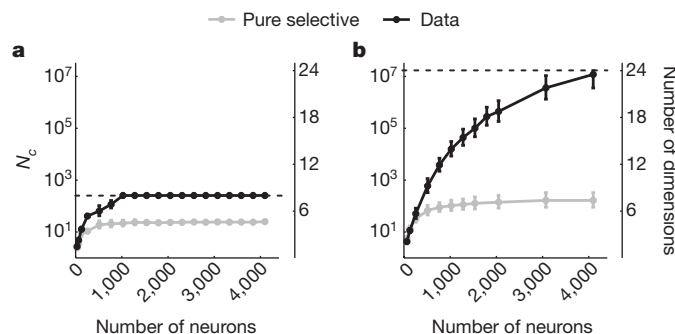
### Neural representations are high-dimensional

To verify that the observed nonlinear mixed selectivity and the diversity of the neural responses are a signature of the high-dimensionality of the neural representations, we set out to quantify the dimensionality of the recorded activity. As this is notoriously difficult in the presence of noise (see Supplementary Section S.10 and ref. 15), we adopted a novel strategy that exploits the relation between dimensionality and the performance of a linear classifier reading out the neural activity (Supplementary Section S.7). Our method relies on the observation that the number of binary classifications that can be implemented by a linear classifier grows exponentially with the number of dimensions of the neural representations of the patterns of activities to be classified (Supplementary Methods M.1). Hence dimensionality can be estimated by counting the number of binary classifications that can be implemented by a linear classifier. The exponential dependence on the dimensionality implies that the number of implementable classifications can vary over several orders of magnitude, allowing for a huge computational advantage of high-dimensional over low-dimensional representations.



In general, a binary linear classifier reads out the activity of the recorded neurons and generates an output that can be either 0 or 1 depending on whether the sum of its weighted inputs is below or above a threshold. The set of weights determines the specific classification that is implemented. In our analysis, the classifier is trained to consistently generate the same output for all recorded trials corresponding to the same experimental condition. The output can be different for the  $c$  different conditions (here,  $c$  equals 24 possible combinations of values for the three task aspects), for a total of  $2^c$  possible binary classifications or, equivalently, mappings from the  $c$  conditions to the binary outputs. Among all possible binary classifications we estimated through cross-validation how many are implementable by a linear classifier of the recorded patterns of activity (Supplementary Methods M.7). The total number of implementable binary classifications  $N_c$  is related to the estimated dimensionality  $d$  of the representations through the expression  $d = \log_2 N_c$ , when the number of inputs is sufficiently large (Supplementary Methods M.1 and Supplementary Section S.7). Accordingly, the dimensionality is bounded by the total number of distinct conditions  $c$ .

Figure 4 shows both  $N_c$  and  $d$  as a function of the number of neurons  $N$  read out by the classifier for two different neural representations. The first neural representation is given by the recorded PFC neurons. For values of  $N$  larger than the number of recorded cells, we computed the performance by introducing additional resampled neurons as previously described for the decoding analysis (Supplementary Methods M.6), after verifying that this does not introduce additional artificial dimensions to the neural representations (Supplementary Section S.8). The second neural representation is generated from simulated pure selectivity neurons that encode only one aspect of the task at a time (Supplementary Methods M.4). We computed  $N_c$  for these neural representations during two task epochs: the one-object (Fig. 4a) and the two-object delay (Fig. 4b). For both epochs,  $N_c$  grows with the size of the neural population  $N$  and it



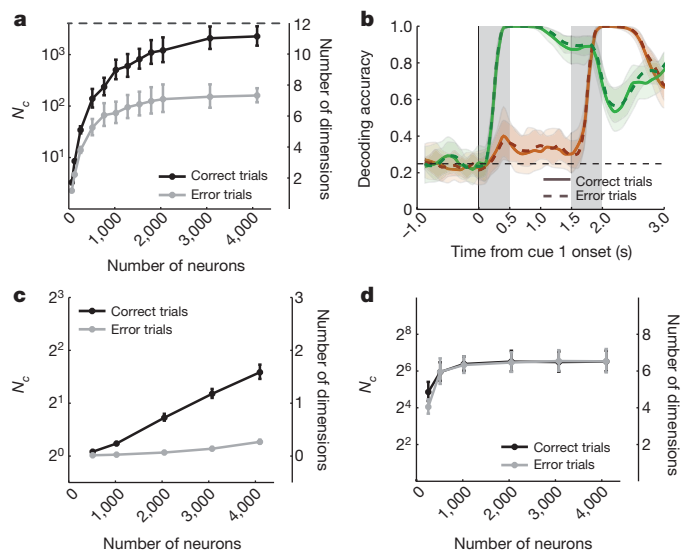
**Figure 4 | The recorded neural representations are high-dimensional.** Number of implementable binary classifications,  $N_c$  (left ordinate, on logarithmic scale) and dimensionality of the inputs (right ordinate, linear scale) for varying number of neurons of the population read out by a linear classifier. **a**, The black trace represents the number  $N_c$  of implementable binary classifications of the vectors of recorded mean firing rates in the 800-ms bin in the middle of the one-object delay period. In this epoch a trial is defined by one of  $c = 8$  different conditions, corresponding to all the combinations of task type and cue 1 objects.  $N_c$  reaches the value that corresponds to the maximal dimensionality  $d = 8$  (indicated by the dashed line). The grey line shows  $N_c$  when the neural representations contain only the responses of artificially generated pure selectivity neurons with a noise level matching that of the data (See Supplementary Methods M.4). **b**, Same plot computed over the 800-ms bin in the middle of the two-object delay period. The advantage of the recorded representations over the pure selectivity neurons is huge. For the recorded data (black line)  $N_c$  reaches  $2^{24}$ , the value that corresponds to the maximal dimensionality  $d = 24$ , given by all possible combinations of cue 1 object, cue 2 object and task type are 24 (dashed line). On the other hand, representations based on pure selectivity (grey line) generate less than 8 dimensions. Error bars are 95% confidence bounds estimated as detailed in Supplementary Methods M.7. See Supplementary Section S.20 for this analysis during the test epochs.

saturates near the value that corresponds to the maximal dimensionality. The asymptotic value of  $N_c$  is always larger for the recorded representations than for the pure selectivity representations. The difference is several orders of magnitude for the two-object delay.

The ability to implement such a large number of classifications is due to the diversity of nonlinear mixed selectivity responses, which often results in seemingly disordered representations and response properties that are not easily interpretable. However, it is important to note that high-dimensional representations could also originate from more ‘orderly’ responses, in which each neuron behaves as a ‘grandmother cell’ that responds only to a single experimental condition—in our case to one out of 24. We ruled out this scenario in our data by verifying that PFC representations are rather dense (Supplementary Section S.11) and that the sparsest responses are not a major contributor to the observed high dimensionality (Supplementary Section S.19).

### Dimensionality predicts animal performance

High-dimensional neural representations encode information that is nonlinearly mixed in a format that is suitable for local neural processing. Therefore high dimensionality could be important for the dynamics of neural circuits that eventually generate behaviour<sup>4–10</sup> (see also Supplementary Section S.1 for an illustration of the computational



**Figure 5 | The dimensionality of the neural representations predicts animal behaviour.** **a**, Same plot as Fig. 4b, with the difference that the analysis is restricted to the recall task and the two curves represent the number of implementable binary classifications of the recorded activity in the case of correct (black) and error (grey) trials. For the correct trials the number of implementable classifications corresponds to a dimensionality that is close to maximal ( $d = 12$ , dashed line). In the error trials the dimensionality drops significantly. **b**, The identity of the two cues can still be decoded in the error trials: decoding accuracy as in Fig. 3 in the correct (continuous lines) and error trials (dashed lines) for the identity of cue 1 (green lines) and cue 2 (orange line). The correct cue identities are perfectly decoded also during error trials. The error bars (shaded areas) are computed as in Fig. 3e–g. **c, d**, Contribution of nonlinear and linear mixed selectivity to the collapse in dimensionality observed in the error trials. **c**, After removing the linear component of mixed selectivity from the response of each neuron, the dimensionality is estimated as in **a**. The dimensionality in the correct trials (black line) is still significantly higher than in the error trials (grey line). **d**, Same as in **c**, but after the nonlinear component of mixed selectivity is subtracted from each neuron. The two curves are not significantly different, indicating that the nonlinear component of mixed selectivity is responsible for the collapse in dimensionality. These analyses were carried out on a subset data set of 121 neurons that were recorded in as many correct trials as error trials during the recall task. **a, c** and **d**, Error bars are 95% confidence bounds estimated as detailed in Supplementary Methods M.7.

advantages of high-dimensional over low-dimensional representations). If this is the case, we should observe a correlation between the dimensionality of the neural representations and the performance of the animal. In particular, the dimensionality should decrease in error trials.

We tested this prediction by analysing the errors made by the monkeys in the recall task (there were too few errors in the recognition task). Figure 5a shows the number of implementable binary classifications and the dimensionality as a function of the number of neurons in the input in the 800-ms time bin in the middle of the two-object delay period, as in Fig. 4b, but only during the recall task and separately for correct and error trials. As predicted, the dimensionality decreases in the error trials. We ruled out that the decrease could be explained by a difference in the number of trials (for the 121 neurons with enough error trials, the difference between the number of correct and error trials is less than 4%), a change in average firing rate (Supplementary Section S.13), in the variability of single neuron activity (Supplementary Section S.14) or in the coding level of the activity (Supplementary Section S.15).

Remarkably, the identity of the two visual cues could still be decoded with high accuracy in the error trials. Figure 5d shows that the cross-validation performances of the decoder of Fig. 3 trained on correct trials and tested on a hold-out set of correct (continuous line) and error trials (dashed lines) are indistinguishable, demonstrating that when the monkeys make a mistake, cue identities are encoded with the same strength as in the correct trials. We verified that this correspondence is not a ceiling effect due to the population decoder saturating at high performance (Supplementary Section S.16).

These results indicate that the collapse in dimensionality observed in the error trials is not due to a failure in coding or remembering the sensory stimuli. Our hypothesis is that nonlinear mixed selectivity, which underlies high dimensionality, is important for the generation of the correct behavioural response. In the error trials this component of the selectivity is disrupted, leading to a collapse in dimensionality, which impairs the ability of downstream readout neurons to produce the correct response.

To test this hypothesis, we quantified the contribution to the collapse in dimensionality of the nonlinear mixed selectivity component of the neuronal response. The nonlinear mixed selectivity component represents the signal that is not described by a linear mixing model (Supplementary Methods M.2). The collapse in dimensionality that predicts the errors of the animal is specifically due to a weakening of the nonlinear component of mixed selectivity. Indeed, the difference in dimensionality between correct and error trials still remains significant after subtracting the linear component of mixed selectivity (Fig. 5c), whereas it is negligible when the nonlinear mixed selectivity component is removed (Fig. 5d). These results are confirmed by a principal component analysis (Supplementary Section S.17).

## Discussion

We showed that the mixed selectivity that is commonly observed in PFC responses can be interpreted as a signature of high-dimensional neural representations. One advantage of high dimensionality is that information about all task-relevant aspects and their combinations is represented in a way that is easily accessible to linear classifiers, such as simple neuron models. The information is distributed across multiple neurons in an 'explicit' format<sup>16</sup> that allows a readout neuron to implement an arbitrary classification of its inputs. Previous studies have already shown that a linear readout is often sufficient to decode particular task aspects or to perform specific tasks (see for example refs 17, 18). Here, by showing that the neural representations are high-dimensional, we demonstrate that any binary choice task involving the 24 experimental conditions that we analysed could be performed by a linear readout.

One of our main results is that the dimensionality of the neural representations collapses in error trials, indicating that nonlinear

mixed selectivity might be important for generating correct behavioural responses. It is tempting to speculate about the causes of this dimensionality collapse. Nonlinear mixed selectivity can change in a way that is compatible with our observations when neurons integrate multiple sources of information, which include those that are relevant for the task and those that are not under experimental control. The change in dimensionality may be caused by the excessive variability of sources that are not task-relevant. In other words, to perform correctly, the brain has to mix nonlinearly the task-relevant sources of information in a way that is consistent across trials. This consistency requires to restrict the contribution of the other sources. This is similar to what has been observed in the premotor cortex, where firing rates tended to be less variable on trials in which the reaction time was shorter<sup>19</sup>. A theoretical argument (Supplementary Section S.18) shows that neurons with a strong nonlinear mixed selectivity are more sensitive than pure selectivity neurons to the task-irrelevant sources of variability. Nonlinear mixed selectivity is most useful but also most fragile. Pure and linear mixed selectivity, which are more robust, make it possible to decode individually all task-relevant aspects even in the error trials, as observed here.

Although high dimensionality is not strictly necessary for generating rich dynamics and performing complex tasks, it is known to greatly simplify the design of local neural circuits<sup>9</sup>. Indeed, realizing a complex and rich dynamics is for some model circuits equivalent to solving a classification problem in which the network has to generate a particular output for each input. In these models this is typically realized by training a subset of neurons to respond in a specific way to an external input or to the internally generated activity. This is equivalent to classifying the activity of the input neurons for every time step. In many situations this activity is read out by downstream circuits. In others it is fed back to the neural circuit to affect its dynamics and hence the statistics of future inputs. Especially in the latter situations, the number of input–output functions or classifications that must be implemented by each neuron can be significantly larger than the number of functions required to simply produce the observed final behavioural response, because the neurons are required to generate the proper output for every time step. For this reason, it is often necessary to expand the dimensionality of the neuronal representations of the external sensory input and the internal state. In recent models<sup>5–10</sup>, the dimensionality of the neuronal representations is expanded by mixing in a nonlinear way the different sources of information in a population of randomly connected neurons. The resulting neuronal representations are high-dimensional (see for example ref. 20), like those observed in PFC, and consistent with high dimensionality, the neurons show mixed selectivity which is diverse across time (that is, in different epochs of the trials) and space (that is, across different neurons). Random connectivity in small brain regions has been suggested on the basis of anatomical reconstructions<sup>21</sup> and recently observed in the connections from the olfactory bulb to the olfactory cortex<sup>22</sup> (see also ref. 14 for a general discussion).

We showed that the recorded mixed selectivity can be useful for the activity to be linearly read out. It is legitimate to ask whether these considerations would still be valid if we consider more complex nonlinear readouts. For example, some of the transformations which increase the dimensionality of the neural representations could be implemented at the level of individual neurons by exploiting dendritic nonlinearities. Our results do not exclude the functional importance of such dendritic processes. They do, however, tend to argue against a scenario where all important nonlinear transformations are carried out at the level of single neurons, a situation where dimensionality expansion could happen in a 'hidden way', and the observable representations provided by the neuronal firing rates could therefore be low-dimensional.

Finally, the particular form of redundancy inherited from high-dimensional representations allows neural circuits to flexibly and quickly adapt to execute new tasks, just as it allows them to implement



arbitrary binary classifications by modifying the weights of a readout neuron (using, for instance, a supervised procedure like the perceptron learning rule<sup>23</sup>). In Supplementary Section S.9 we show an example of this flexibility by training a simulated neuron to perform a new virtual task based on the recorded activity. High dimensionality might therefore be at the basis of the mechanisms underlying the remarkable adaptability of the neural coding observed in the PFC<sup>13</sup> and, as such, be an important element to answer fundamental questions that try to map cognitive to neurophysiological functions.

In conclusion, the measured dimensionality of the neural representations in PFC is high, and errors follow a collapse in dimensionality. This provides us with a motivation to shift the focus of attention from pure selectivity neurons, which are easily interpretable, to the widely observed but rarely analysed mixed selectivity neurons, especially in the complex task designs that are becoming progressively more accessible to investigation.

## METHODS SUMMARY

The formal definitions of dimensionality and mixed selectivity are in Supplementary Methods M.1 and M.2, respectively. The procedures for removing selectivity, decoding task-relevant aspects and resampling neurons used in Fig. 3 are explained in Supplementary Methods M.3, M.5 and M.6. The dimensionality estimate of Fig. 4 is detailed in Supplementary Methods M.7. The analysis of the linear and nonlinear components of Fig. 5 is in Supplementary Methods M.8.

Received 11 September 2012; accepted 5 April 2013.

Published online 19 May 2013.

1. Asaad, W. F., Rainer, G. & Miller, E. K. Neural activity in the primate prefrontal cortex during associative learning. *Neuron* **21**, 1399–1407 (1998).
2. Mansouri, F. A., Matsumoto, K. & Tanaka, K. Prefrontal cell activities related to monkeys' success and failure in adapting to rule changes in a Wisconsin card sorting test analog. *J. Neurosci.* **26**, 2745–2756 (2006).
3. Warden, M. R. & Miller, E. K. Task-dependent changes in short-term memory in the prefrontal cortex. *J. Neurosci.* **30**, 15801–15810 (2010).
4. Buonomano, D. V. & Merzenich, M. M. Temporal information transformed into a spatial code by a neural network with realistic properties. *Science* **267**, 1028–1030 (1995).
5. Maass, W., Natschlager, T. & Markram, H. Real-time computing without stable states: a new framework for neural computation based on perturbations. *Neural Comput.* **14**, 2531–2560 (2002).
6. Jaeger, H. & Haas, H. Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication. *Science* **304**, 78–80 (2004).
7. Buonomano, D. V. & Maass, W. State-dependent computations: spatiotemporal processing in cortical networks. *Nature Rev. Neurosci.* **10**, 113–125 (2009).
8. Sussillo, D. & Abbott, L. F. Generating coherent patterns of activity from chaotic neural networks. *Neuron* **63**, 544–557 (2009).
9. Rigotti, M., Ben Dayan Rubin, D. D., Wang, X.-J. & Fusi, S. Internal representation of task rules by recurrent dynamics: the importance of the diversity of neural responses. *Front. Comput. Neurosci.* **4**, 24 (2010).
10. Pascanu, R. & Jaeger, H. A neurodynamical model for working memory. *Neural Netw.* **24**, 199–207 (2011).
11. Cortes, C. & Vapnik, V. Support-vector networks. *Mach. Learn.* **20**, 273–297 (1995).
12. Warden, M. R. & Miller, E. K. The representation of multiple objects in prefrontal neuronal delay activity. *Cereb. Cortex* **17** (Suppl. 1), i41–i50 (2007).
13. Duncan, J. An adaptive coding model of neural function in prefrontal cortex. *Nature Rev. Neurosci.* **2**, 820–829 (2001).
14. Yuste, R. Dendritic spines and distributed circuits. *Neuron* **71**, 772–781 (2011).
15. Machens, C. K., Romo, R. & Brody, C. D. Functional, but not anatomical, separation of “what” and “when” in prefrontal cortex. *J. Neurosci.* **30**, 350–360 (2010).
16. DiCarlo, J. J., Zoccolan, D. & Rust, N. C. How does the brain solve visual object recognition? *Neuron* **73**, 415–434 (2012).
17. Meyers, E. M., Freedman, D. J., Kreiman, G., Miller, E. K. & Poggio, T. Dynamic population coding of category information in inferior temporal and prefrontal cortex. *J. Neurophysiol.* **100**, 1407–1419 (2008).
18. Klampfl, S., David, S. V., Yin, P., Shamma, S. A. & Maass, W. A quantitative analysis of information about past and present stimuli encoded by spikes of A1 neurons. *J. Neurophysiol.* **108**, 1366–1380 (2012).
19. Churchland, M. M., Yu, B. M., Ryu, S. I., Santhanam, G. & Shenoy, K. V. Neural variability in premotor cortex provides a signature of motor preparation. *J. Neurosci.* **26**, 3697–3712 (2006).
20. Barak, O., Rigotti, M. & Fusi, S. The sparseness of mixed selectivity neurons controls the generalization–discrimination trade-off. *J. Neurosci.* **33**, 3844–3856 (2013).
21. Braitenberg, V. & Schüz, A. *Cortex: Statistics and Geometry of Neuronal Connectivity* 2nd edn (Springer, 1998).
22. Sosulski, D. L., Bloom, M. L., Cutforth, T., Axel, R. & Datta, S. R. Distinct representations of olfactory information in different cortical centres. *Nature* **472**, 213–216 (2011).
23. Rosenblatt, F. *Principles of Neurodynamics* (Spartan Books, 1962).

Supplementary Information is available in the online version of the paper.

**Acknowledgements** We are grateful to L.F. Abbott for comments on the manuscript and for discussions. Work supported by the Gatsby Foundation, the Swartz Foundation and the Kavli Foundation. M.R. is supported by Swiss National Science Foundation grant PBSKP3-133357 and the Janggen-Poehn Foundation; N.D.D. is supported by the McKnight Foundation and the McDonnell Foundation; E.K.M. is supported by NIMH grant 5-R37-MH087027-04 and The Picower Foundation; M.R.W. from the Brain & Behavior Research Foundation and the NARSAD Young Investigator grant.

**Author Contributions** M.R.W. and E.K.M. performed the experiments and collected the data. M.R., O.B., X.-J.W. and S.F. developed the theoretical framework. M.R., O.B., N.D.D. and S.F. conceived the data analyses. M.R. performed the data analyses. M.R., O.B. and S.F. wrote the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.F. (sf2237@columbia.edu).

# An anti-glitch in a magnetar

R. F. Archibald<sup>1</sup>, V. M. Kaspi<sup>1</sup>, C. -Y. Ng<sup>1,2</sup>, K. N. Gourgouliatos<sup>1</sup>, D. Tsang<sup>1</sup>, P. Scholz<sup>1</sup>, A. P. Beardmore<sup>3</sup>, N. Gehrels<sup>4</sup> & J. A. Kennea<sup>5</sup>

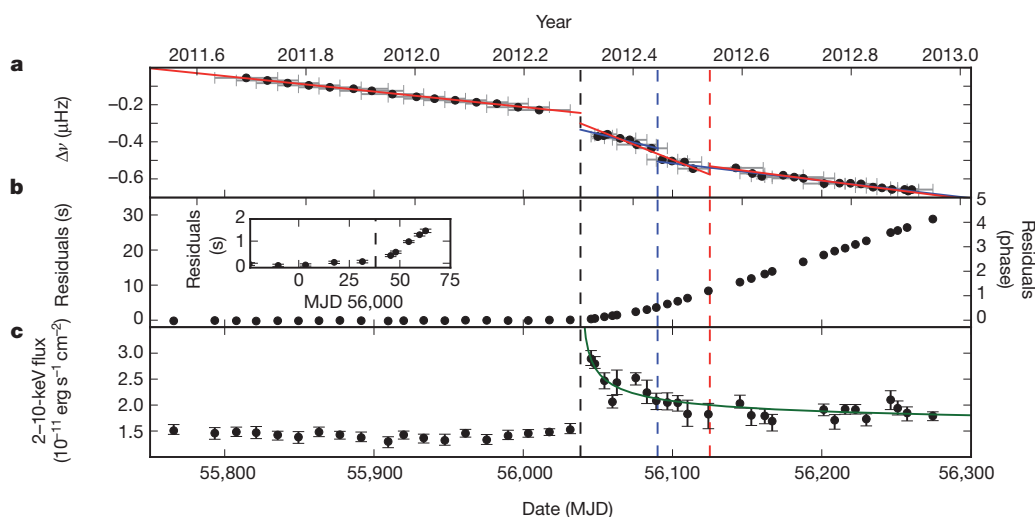
Magnetars are neutron stars with X-ray and soft  $\gamma$ -ray outbursts thought to be powered by intense internal magnetic fields<sup>1</sup>. Like conventional neutron stars in the form of radio pulsars, magnetars exhibit ‘glitches’ during which angular momentum is believed to be transferred between the solid outer crust and the superfluid component of the inner crust<sup>2–4</sup>. The several hundred observed glitches in radio pulsars<sup>5,6</sup> and magnetars<sup>7</sup> have involved a sudden spin-up (increase in the angular velocity) of the star, presumably because the interior superfluid was rotating faster than the crust. Here we report X-ray timing observations of the magnetar 1E 2259+586 (ref. 8), which exhibited a clear ‘anti-glitch’—a sudden spin-down. We show that this event, like some previous magnetar spin-up glitches<sup>9</sup>, was accompanied by multiple X-ray radiative changes and a significant spin-down rate change. Such behaviour is not predicted by models of neutron star spin-down and, if of internal origin, is suggestive of differential rotation in the magnetar, supporting the need for a rethinking of glitch theory for all neutron stars<sup>10,11</sup>.

1E 2259+586 is a magnetar with a rotation period of about 7 s, with a spin-inferred surface dipolar magnetic field strength of  $5.9 \times 10^{13}$  G. Over 16 years of monitoring with the Rossi X-ray Timing Explorer, 1E 2259+586 has shown a very stable spin-down rate and pulsed flux, with the exception of two spin-up glitches in 2002 (ref. 9) and 2007 (ref. 12), and a small timing event in 2009 (ref. 12). The 2002 glitch was also accompanied by an increase in X-ray luminosity by a factor of 20 (ref. 9) and X-ray bursts<sup>13</sup>, neither of which was seen in the 2007 glitch.

We began monitoring 1E 2259+586 with NASA’s Swift X-ray Telescope<sup>14</sup> in July 2011. Observations were made every 2–3 weeks, with typical exposure times of 4 ks. From each observation, we obtained a pulse time-of-arrival (TOA) by folding the X-ray time series at the current pulse period and aligning this folded light curve with a high signal-to-noise template.

We fitted the pulse TOAs to a long-term timing model that keeps track of every rotation of the neutron star. This model predicts when the pulses should arrive at Earth, taking into account the pulsar rotation as well as astrometric terms. We compared the observed TOAs with the model predictions, and obtained best-fit parameters by  $\chi^2$  minimization, using the TEMPO2<sup>15</sup> software package. Until the observation on 14 April 2012 (modified Julian date, MJD 56,031.18), these TOAs were well fitted using only a frequency and first frequency derivative, as shown in Fig. 1.

The subsequent data, however, clearly were not predicted by this simple model. TOAs starting on 28 April 2012 (MJD 56,045.01) showed an apparently instantaneous change of the frequency—which we dub an ‘anti-glitch’. On 21 April 2012 (MJD 56,038), consistent with the epoch of this sudden spin-down, a 36-ms hard X-ray burst was detected by the Fermi Gamma-ray Burst Monitor<sup>16</sup>, with a fluence of about  $6 \times 10^{-8}$  erg cm<sup>-2</sup> in the 10–1,000 keV range. No untriggered bursts from the Gamma-ray Burst Monitor were seen within three days of the observed burst<sup>16</sup>. Also, on 28 April 2012 (MJD 56,045.01), coincident with the nearest post-anti-glitch observation, we detected an increase in the 2–10-keV flux by a factor of  $2.00 \pm 0.09$ .



**Figure 1 | Timing and flux properties of 1E 2259+586 around the 2012 event.** **a**, 1E 2259+586’s spin frequency as a function of time, determined by short-term fitting of (typically) five TOAs. The grey horizontal error bars indicate the ranges of dates used to fit the frequency, and the vertical error bars (generally smaller than the points) are standard  $1\sigma$  uncertainties. The red and blue solid lines in **a** represent the fits to the pulse TOAs, as displayed in Table 1, with red representing model 1, and blue model 2. **b**, Timing residuals (differences between the initial model and observed data) of 1E 2259+586 after

fitting only for the pre-anti-glitch timing solution. The inset shows the same timing residuals, zooming in on the anti-glitch epoch. **c**, The absorbed 2–10-keV X-ray flux. The error bars indicate the  $1\sigma$  uncertainties, and the green line is the best-fit power-law decay curve with an index of  $-0.38 \pm 0.04$ . The dashed vertical lines running through all panels indicate the glitch epochs, the black line being the anti-glitch, and blue and red lines the second event in the models shown in Table 1. The timing residuals for these fits can be seen in the Supplementary Information.

<sup>1</sup>Department of Physics, McGill University, Montreal, Quebec H3A 2T8, Canada. <sup>2</sup>Department of Physics, The University of Hong Kong, Pokfulam Road, Hong Kong. <sup>3</sup>Department of Physics and Astronomy, University of Leicester, University Road, Leicester LE1 7RH, UK. <sup>4</sup>Astrophysics Science Division, NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA. <sup>5</sup>Department of Astronomy and Astrophysics, 525 Davey Laboratory, Pennsylvania State University, University Park, Pennsylvania 16802, USA.



(see Fig. 1). The 2–10-keV flux increase was also accompanied by a change in the hardness ratio, defined as the ratio of the 4–10 keV flux to the 2–4 keV fluxes, from  $0.10 \pm 0.02$  to  $0.18 \pm 0.02$ . This flux increase subsequently decayed, following a power-law model with  $\alpha = -0.38 \pm 0.04$  (see Fig. 1). The flux increase was accompanied by a moderate change in the pulse profile: the addition of a sinusoid centred between the usual two peaks in the pulse profile. This modified pulse profile relaxed back to the usual shape on a timescale similar to that of the flux. We verified that this profile change did not affect the TOAs determined near the anti-glitch epoch.

This remarkable spin-down event was immediately followed by an extended period of enhanced spin-down rate. This anti-glitch and spin-down rate change can be modelled well by an instantaneous change in the frequency and frequency derivative, followed by a second sudden event. We have found two possible timing models to describe the pulsar's behaviour, described in full in Table 1. In the first, there is an instantaneous change in frequency and frequency derivative by  $\Delta\nu = -4.5(6) \times 10^{-8}$  Hz (where  $\Delta\nu/\nu = -3.1(4) \times 10^{-7}$ ) and  $\Delta\dot{\nu} = -2.7(2) \times 10^{-14}$  Hz s $^{-1}$  on 18 April (MJD 56,035(2)). This enhanced spin-down episode ended with a second glitch, this time a spin-up event, of amplitude  $\Delta\nu = 3.6(7) \times 10^{-8}$  Hz ( $\Delta\nu/\nu = 2.6(5) \times 10^{-7}$ ) and  $\Delta\dot{\nu} = 2.6(2) \times 10^{-14}$  Hz s $^{-1}$ .

In the second model, the spin evolution can be described by two anti-glitches, instead of an anti-glitch/glitch pair. In this model, a change of  $\Delta\nu = -9(1) \times 10^{-8}$  Hz ( $\Delta\nu/\nu = -6.3(7) \times 10^{-7}$ ) and  $\Delta\dot{\nu} = -1.3(4) \times 10^{-14}$  Hz s $^{-1}$  occurred on 21 April (MJD 56,038(2)). This period ended with a second anti-glitch of amplitude  $\Delta\nu = -6.8(8) \times 10^{-8}$  Hz ( $\Delta\nu/\nu = 4.8(5) \times 10^{-7}$ ) and  $\Delta\dot{\nu} = 1.1(4) \times 10^{-14}$  Hz s $^{-1}$ .

The full timing parameters for both possible models are presented in Table 1. Note that neither model is preferred on statistical grounds; however, models involving a single initial anti-glitch and subsequent relaxation with no second impulsive event are ruled out to high confidence. We also note that no significant radiative, or profile, changes can be associated with either of the possible second impulsive events.

In either model a sudden spin-down at the epoch of the Fermi burst is unambiguously required to model the observed TOAs properly. Although the amplitude of this anti-glitch in either model is not unusual, the fact that it is a sudden spin-down is remarkable. The net effect of this active period are changes to the spin frequency and its first derivative  $\Delta\nu = -2.06(8) \times 10^{-7}$  Hz ( $\Delta\nu/\nu = -1.44(6) \times 10^{-6}$ ) and a  $\Delta\dot{\nu} = -1.3(4) \times 10^{-15}$  Hz s $^{-1}$ .

**Table 1 | Timing parameters for 1E 2259+586**

Parameter	Value
Observation dates	23 July 2011 to 1 January 2013
Dates (MJD)	55,765.829 to 56,293.332
Epoch (MJD)	55,380.000
Number of TOAs	51
$\nu$ (s $^{-1}$ )	0.143, 285, 110(4)
$\dot{\nu}$ (s $^{-2}$ )	$-9.80(9) \times 10^{-15}$
<b>Model 1</b>	
Glitch epoch 1 (MJD)	56,035(2)
$\Delta\nu_1$ (s $^{-1}$ )	$-4.5(6) \times 10^{-8}$
$\Delta\dot{\nu}_1$ (s $^{-2}$ )	$-2.7(2) \times 10^{-14}$
Glitch epoch 2 (MJD)	56,125(2)
$\Delta\nu_2$ (s $^{-1}$ )	$3.6(7) \times 10^{-8}$
$\Delta\dot{\nu}_2$ (s $^{-2}$ )	$2.6(2) \times 10^{-14}$
Root-mean-square residuals (ms)	56.3
$\chi^2/\nu$	45.4/44
<b>Model 2</b>	
Glitch epoch 1 (MJD)	56,039(2)
$\Delta\nu_1$ (s $^{-1}$ )	$-9(1) \times 10^{-8}$
$\Delta\dot{\nu}_1$ (s $^{-2}$ )	$-1.3(4) \times 10^{-14}$
Glitch epoch 2 (MJD)	56,090(3)
$\Delta\nu_2$ (s $^{-1}$ )	$-6.8(8) \times 10^{-8}$
$\Delta\dot{\nu}_2$ (s $^{-2}$ )	$1.1(4) \times 10^{-14}$
Root-mean-square residuals (ms)	51.5
$\chi^2/\nu$	38.1/44

Numbers in parentheses are TEMPO2-reported  $1\sigma$  uncertainties.

Sudden spin-down glitches have not previously been observationally demonstrated, though some magnetar events have been suggestive. A spin-down in magnetar SGR 1900+14 (ref. 17) occurred during an 80-day gap in the source monitoring, but could have been a gradual slowdown, as was also possible for the 2009 timing event in 1E 2259+586 (ref. 12). Net spin-downs have been seen in magnetar 4U 0142+61 (ref. 18) and in the high-magnetic-field-rotation-powered pulsar PSR J1846–0258 (ref. 19) but were due to spin-up glitch over-recoveries on timescales of 17 and 127 days, respectively. If the 1E 2259+586 event were due to a spin-up glitch and subsequent over-recovery, we place a  $3\sigma$  upper limit on the recovery decay time of 3.9 days for a spin-up of size  $\Delta\nu/\nu = 1 \times 10^{-6}$ . Even for an infinitesimally small spin-up glitch, the decay time was less than 6.6 days, far shorter than any previously observed magnetar recovery timescales.

Following the detection of the anti-glitch, we looked for evidence of particle outflow, proposed as a possible mechanism for the apparent spin-down in SGR 1900+14 (ref. 20). We carried out radio imaging on 21 August 2012 using the Expanded Very Large Array (New Mexico, USA) in the B-array configuration with a 240-min integration time. This yielded images with effective angular resolution of  $1.2''$ . We performed standard flagging, calibration and imaging using the Common Astronomy Software Applications (CASA) package<sup>21</sup>. No source was found at the position of 1E 2259+586, and we place a  $3\sigma$  flux density limit of  $7.2 \mu\text{Jy}$  at 7 GHz for a point source. This is significantly lower than the previous upper limit of  $50 \mu\text{Jy}$  at 1.4 GHz (ref. 9). If a putative outflow were expanding at  $0.7c$  (where  $c$  is the velocity of light in a vacuum), as was the case for a radio outflow from SGR 1806–20 (ref. 22) at the time of its outburst, we would expect a nebular radius of  $4''$ . For this radius, we obtain a  $3\sigma$  flux density limit of  $0.46 \text{ mJy}$ . The limit is more stringent if the size is smaller, and reduces to  $7.2 \mu\text{Jy}$  if unresolved.

In X-rays, we also detected no evidence for such outflow in a 10-ks Chandra High Resolution Camera (HRC)-I image taken on 21 August 2012. Using simulations, we place an upper limit on X-ray flux from a putative outflow at 2% of the total 1–10-keV X-ray emission of the magnetar, for a  $4''$  circular nebula with a Crab-like spectrum.

There are two main possibilities for the origin of the anti-glitch: either an internal or external mechanism, as follows. An impulse-like angular-momentum transfer between regions of more slowly spinning superfluid and the crust could be the source of the anti-glitch<sup>20</sup>. A slower angular-momentum transfer to such a region or the decoupling of a significant amount of the moment of inertia of the star could account for the enhanced spin-down rate. The second event, either glitch or anti-glitch, can similarly be modelled by angular-momentum transfers from differentially rotating regions of the neutron star superfluid. The radiatively quiet nature of the second event does not pose a problem for the internal model because many glitches are observed to be radiatively silent<sup>7</sup>. The behaviour indicated by an impulsive anti-glitch offers new evidence for possible significant internal structural changes and differential rotation in magnetars at glitch epochs.

An external model such as an outflow along the open field lines of the magnetosphere<sup>20,23,24</sup>, or a sudden twisting of the field lines<sup>25</sup>, could be the cause of the anomalous spin-down behaviour. However, in a wind model, the second timing event should also be accompanied by a radiative change, like the first one. If this behaviour was caused by twisting magnetic field lines, it should be followed by a gradual untwisting and a similar behaviour reflected in  $\dot{\nu}$  (ref. 26; see Supplementary Information). We suggest that this magnetar anti-glitch, X-ray outburst, and subsequent evolution indicate the need for a rethinking of glitch theory for all neutron stars<sup>10,11</sup>.

Received 7 February; accepted 4 April 2013.

Published online 15 May 2013.

1. Thompson, C. & Duncan, R. C. The soft gamma repeaters as very strongly magnetized neutron stars—I. Radiative mechanism for outbursts. *Mon. Not. R. Astron. Soc.* **275**, 255–300 (1995).
2. Anderson, P. W. & Itoh, N. Pulsar glitches and restlessness as a hard superfluidity phenomenon. *Nature* **256**, 25–27 (1975).

3. Pines, D. & Alpar, M. A. Superfluidity in neutron stars. *Nature* **316**, 27–32 (1985).
4. Link, B., Epstein, R. I. & van Riper, K. A. Pulsar glitches as probes of neutron star interiors. *Nature* **359**, 616–618 (1992).
5. Espinoza, C. M., Lyne, A. G., Stappers, B. W. & Kramer, M. A study of 315 glitches in the rotation of 102 pulsars. *Mon. Not. R. Astron. Soc.* **414**, 1679–1704 (2011).
6. Yu, M. *et al.* Detection of 107 glitches in 36 southern pulsars. *Mon. Not. R. Astron. Soc.* **429**, 688–724 (2013).
7. Dib, R., Kaspi, V. M. & Gavril, F. P. Glitches in anomalous X-ray pulsars. *Astrophys. J.* **673**, 1044–1061 (2008).
8. Fahlman, G. G. & Gregory, P. C. An X-ray pulsar in SNR G109.1–1.0. *Nature* **293**, 202–204 (1981).
9. Kaspi, V. M. *et al.* A major soft gamma repeater-like outburst and rotation glitch in the no-longer-so-anomalous X-ray pulsar 1E 2259+586. *Astrophys. J.* **588**, L93–L96 (2003).
10. Ruderman, M., Zhu, T. & Chen, K. Neutron star magnetic field evolution, crust movement, and glitches. *Astrophys. J.* **492**, 267 (1998).
11. Andersson, N., Glampedakis, K., Ho, W. C. G. & Espinoza, C. M. Pulsar glitches: the crust is not enough. *Phys. Rev. Lett.* **109**, 241103 (2012).
12. İçdem, B., Baykal, A., Inam, S. & Ç. RXTE timing analysis of the anomalous X-ray pulsar 1E 2259+586. *Mon. Not. R. Astron. Soc.* **419**, 3109–3114 (2012).
13. Gavril, F. P., Kaspi, V. M. & Woods, P. M. A comprehensive study of the X-ray bursts from the magnetar candidate 1E 2259+586. *Astrophys. J.* **607**, 959–969 (2004).
14. Burrows, D. *et al.* The Swift X-ray telescope. *Space Sci. Rev.* **120**, 165–195 (2005).
15. Hobbs, G. B., Edwards, R. T. & Manchester, R. N. TEMPO2, a new pulsar-timing package—I. An overview. *Mon. Not. R. Astron. Soc.* **369**, 655–672 (2006).
16. Foley, S., Kouveliotou, C., Kaneko, Y. & Collazzi, A. Fermi/GBM detection of a burst from the magnetar 1E 2259+5. *GRB Coord. Netw.* **13280**, 1 (2012).
17. Woods, P. M. *et al.* Variable spin-down in the soft gamma repeater SGR 1900+14 and correlations with burst activity. *Astrophys. J.* **524**, L55–L58 (1999).
18. Gavril, F. P., Dib, R. & Kaspi, V. M. The 2006–2007 active phase of anomalous X-ray pulsar 4U 0142+61: radiative and timing changes, bursts, and burst spectral features. *Astrophys. J.* **736**, 138 (2011).
19. Livingstone, M. A., Kaspi, V. M. & Gavril, F. P. Timing behavior of the magnetically active rotation-powered pulsar in the supernova remnant Kesteven 75. *Astrophys. J.* **710**, 1710–1717 (2010).
20. Thompson, C. *et al.* Physical mechanisms for the variable spin-down and light curve of SGR 1900+14. *Astrophys. J.* **543**, 340–350 (2000).
21. International Consortium Of Scientists. CASA: Common Astronomy Software Applications. Astrophysics Source Code Library 7013 (2011); <http://casa.nrao.edu/index.shtml>.
22. Granot, J. *et al.* Diagnosing the outflow from the SGR 1806–20 giant flare with radio observations. *Astrophys. J.* **638**, 391–396 (2006).
23. Thompson, C. & Blaes, O. Magnetohydrodynamics in the extreme relativistic limit. *Phys. Rev. D* **57**, 3219–3234 (1998).
24. Harding, A. K., Contopoulos, I. & Kazanas, D. Magnetar spin-down. *Astrophys. J.* **525**, L125–L128 (1999).
25. Beloborodov, A. M. Untwisting magnetospheres of neutron stars. *Astrophys. J.* **703**, 1044–1060 (2009).
26. Parfrey, K., Beloborodov, A. M. & Hui, L. Twisting, reconnecting magnetospheres and magnetar spindown. *Astrophys. J.* **754**, L12 (2012).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** V.M.K. acknowledges support from the Natural Sciences and Engineering Research Council of Canada Discovery Grant and the John C. Polanyi Award, from the Canadian Institute for Advanced Research, from Fonds de Recherche Nature et Technologies Québec, from the Canada Research Chairs Program, and from the Lorne Trottier Chair in Astrophysics and Cosmology. D.T. was supported by the Lorne Trottier Chair in Astrophysics and Cosmology and the Canadian Institute for Advanced Research. K.N.G. was supported by the Centre de Recherche en Astrophysique du Québec. We thank H. Medlin and J. Gelfand for help with the EVLA observation. We thank D. Eichler, B. Link, M. Lyutikov and C. Thompson for useful discussions. We acknowledge the use of public data from the Swift data archive.

**Author Contributions** R.F.A. performed the data analysis and wrote portions of the analysis software. V.M.K. designed the study, was the project leader for the Swift data, proposed for the Chandra data and assisted with the interpretation of the data analysis and the theoretical implications. C.Y.N. proposed for the VLA data and reduced them and the Chandra data. K.N.G. and D.T. assisted with the theoretical implications. P.S. wrote significant portions of the Swift analysis software. A.P.B., N.G. and J.A.K. assisted with Swift observations and data analysis. R.F.A. wrote the paper with guidance from V.M.K. and with significant input from all co-authors.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to V.M.K. ([vkaspi@physics.mcgill.ca](mailto:vkaspi@physics.mcgill.ca)).



# Cloning of Dirac fermions in graphene superlattices

L. A. Ponomarenko<sup>1</sup>, R. V. Gorbachev<sup>2</sup>, G. L. Yu<sup>1</sup>, D. C. Elias<sup>1</sup>, R. Jalil<sup>2</sup>, A. A. Patel<sup>3</sup>, A. Mishchenko<sup>1</sup>, A. S. Mayorov<sup>1</sup>, C. R. Woods<sup>1</sup>, J. R. Wallbank<sup>3</sup>, M. Mucha-Kruczynski<sup>3</sup>, B. A. Piot<sup>4</sup>, M. Potemski<sup>4</sup>, I. V. Grigorieva<sup>1</sup>, K. S. Novoselov<sup>1</sup>, F. Guinea<sup>5</sup>, V. I. Fal'ko<sup>3</sup> & A. K. Geim<sup>1,2</sup>

Superlattices have attracted great interest because their use may make it possible to modify the spectra of two-dimensional electron systems and, ultimately, create materials with tailored electronic properties<sup>1–8</sup>. In previous studies (see, for example, refs 1–8), it proved difficult to realize superlattices with short periodicities and weak disorder, and most of their observed features could be explained in terms of cyclotron orbits commensurate with the superlattice<sup>1–4</sup>. Evidence for the formation of superlattice minibands (forming a fractal spectrum known as Hofstadter's butterfly<sup>9</sup>) has been limited to the observation of new low-field oscillations<sup>5</sup> and an internal structure within Landau levels<sup>6–8</sup>. Here we report transport properties of graphene placed on a boron nitride substrate and accurately aligned along its crystallographic directions. The substrate's moiré potential<sup>10–12</sup> acts as a superlattice and leads to profound changes in the graphene's electronic spectrum. Second-generation Dirac points<sup>13–22</sup> appear as pronounced peaks in resistivity, accompanied by reversal of the Hall effect. The latter indicates that the effective sign of the charge carriers changes within graphene's conduction and valence bands. Strong magnetic fields lead to Zak-type cloning<sup>23</sup> of the third generation of Dirac points, which are observed as numerous neutrality points in fields where a unit fraction of the flux quantum pierces the superlattice unit cell. Graphene superlattices such as this one provide a way of studying the rich physics expected in incommensurable quantum systems<sup>7–9,22–24</sup> and illustrate the possibility of controllably modifying the electronic spectra of two-dimensional atomic crystals by varying their crystallographic alignment within van der Waals heterostructures<sup>25</sup>.

Since the first observation of Weiss oscillations<sup>1,2</sup>, two-dimensional electronic systems subjected to a periodic potential have been studied in great detail<sup>3–8</sup>. The advent of graphene rapidly sparked interest in graphene superlattices<sup>13–22</sup>. The principal novelty of such superlattices is the Dirac-like spectrum and the fact that charge carriers are not buried deep under the surface, which allows a relatively strong superlattice potential on the nanometre scale. One promising method of making nanoscale graphene superlattices is the use of a potential induced by another crystal. For example, graphene placed on top of graphite or hexagonal boron nitride (hBN) exhibits a moiré pattern<sup>10–12,26</sup>, and the graphene's tunnelling density of states becomes strongly modified<sup>12,26</sup>, indicating the formation of superlattice minibands. This spectral reconstruction occurs near the edges of the superlattice Brillouin zone (SBZ) that are characterized<sup>12,22</sup> by wavevector  $G = 4\pi/\sqrt{3}D$  and energy  $E_S = \hbar v_F G/2$  ( $D$  is the superlattice period,  $v_F$  is graphene's Fermi velocity and  $\hbar$  is Planck's constant divided by  $2\pi$ ).

To observe moiré minibands in transport properties, graphene has to be doped so that the Fermi energy reaches the reconstructed part of the spectrum. This imposes severe constraints on the misalignment angle,  $\theta$ , of the graphene relative to the hBN substrate. Indeed,  $D$  is determined from  $\theta$  and the 1.8% difference between the two lattice constants<sup>12</sup>. In the case of perfect alignment ( $\theta = 0$ ),  $D$  has a maximum value of  $\sim 13 \pm 1$  nm (ref. 12), which yields  $E_S \approx 0.2$  eV. This energy

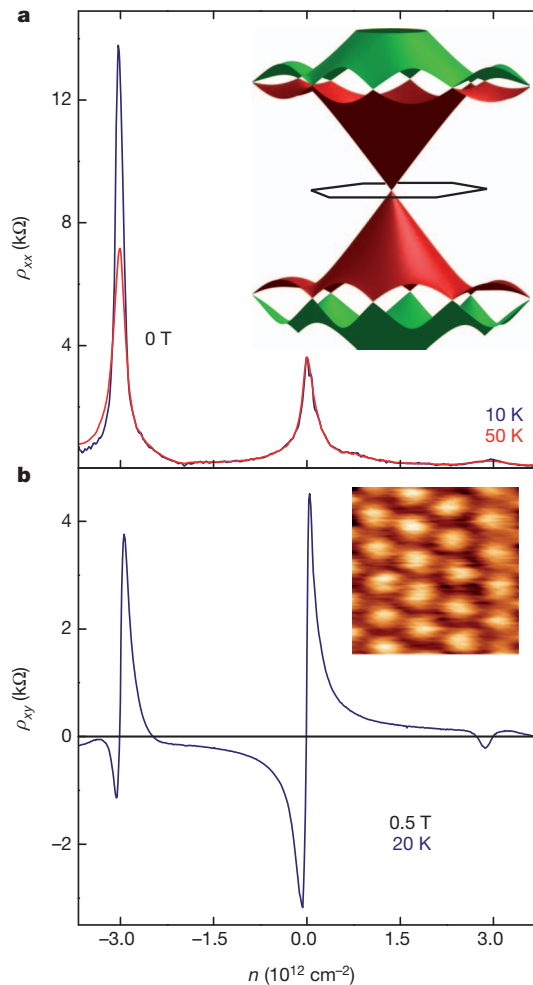
scale corresponds to a carrier density of  $n \approx 3 \times 10^{12} \text{ cm}^{-2}$ , which is achievable by field-effect doping. However, misorientation by only  $2^\circ$  decreases  $D$  by a factor of two<sup>12</sup>, and fourfold greater values of  $n$  are necessary for the Fermi energy to reach the edges of the first SBZ. In practice, studies of the superlattice spectrum in monolayer graphene require  $\theta \leq 1^\circ$  (Methods).

Here we study high-mobility encapsulated graphene devices that are similar to those reported previously<sup>27</sup> but which involve a new element: crystallographic alignment between the graphene and the hBN with a precision of  $\sim 1^\circ$ . Figure 1 shows typical behaviour of longitudinal and Hall resistivities ( $\rho_{xx}$  and  $\rho_{xy}$ , respectively) for our aligned devices. There is the standard peak in  $\rho_{xx}$  at  $n = 0$ , graphene's main neutrality point. In addition, two other peaks appear symmetrically, one on either side of the main neutrality point, at high doping,  $n = \pm n_S$ . At low temperatures ( $T$ ), the secondary peak on the hole side is stronger than that at the main neutrality point, whereas that on the electron side is  $\sim 10$  times weaker. The reversal in sign of  $\rho_{xy}$  (Fig. 1b) cannot be explained by additional scattering and proves that hole-like and electron-like carriers appear in the conduction and, respectively, valence bands of graphene. We attribute the extra neutrality points to the superlattice potential induced by the hBN, which results in minibands featuring isolated secondary Dirac points (Fig. 1a, inset). This interpretation agrees with theory<sup>12–22</sup> and the tunnelling features reported in ref. 12, including the fact that those were stronger in the valence band.

Near the main neutrality point, the aligned devices have transport characteristics typical for graphene on hBN<sup>27,28</sup>. The conductivity  $\sigma(n) = 1/\rho_{xx}$  varies linearly with  $n$  and can therefore be described in terms of constant mobility,  $\mu$ . For the reported devices, we find that  $\mu \approx (20\text{--}80) \times 10^3 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$  for  $|n| > 10^{11} \text{ cm}^{-2}$ . Around the secondary neutrality points,  $\sigma$  depends linearly on  $n - n_S$ . At the hole-side secondary neutrality point (hSNP), at low temperature  $\mu$  is practically the same as at the main neutrality point, whereas near the electron-side secondary neutrality point we find even higher values,  $\mu \approx (30\text{--}100) \times 10^3 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ . However, at the main and secondary neutrality points the  $T$  dependences of both  $\mu$  and the minimum conductivities are very different. This is discussed in Supplementary Information, section 1, and here we note only that the observed functions  $\sigma(T)$  do not support the idea of major energy gaps being induced by the superlattice at the cloned secondary Dirac points<sup>19–22</sup> (Fig. 1a, inset). Furthermore, following the approach described in ref. 29, we analysed the thermal broadening of the peaks in  $\rho_{xx}$  (Supplementary Information, section 2). The analysis proves that the spectrum at the secondary neutrality points is linear, that is, Dirac-like, in agreement with theory<sup>13–22</sup>.

Figure 2 shows the evolution of  $\rho_{xx}(n)$  with increasing perpendicular magnetic field,  $B$ . Near the main Dirac point, we observe the standard<sup>30</sup> quantum Hall effect (QHE) for graphene, with plateaux in  $\rho_{xy}$  and zeros in  $\rho_{xx}$  at filling factors  $\nu \equiv n\phi_0/B = \pm 2, \pm 6, \pm 10, \dots$  where  $\phi_0$  is the flux quantum. Fan diagrams around the secondary Dirac points are different (Fig. 2). The resistance peak of the hSNP first broadens with increasing  $B$  and then splits into two maxima. The maxima correspond

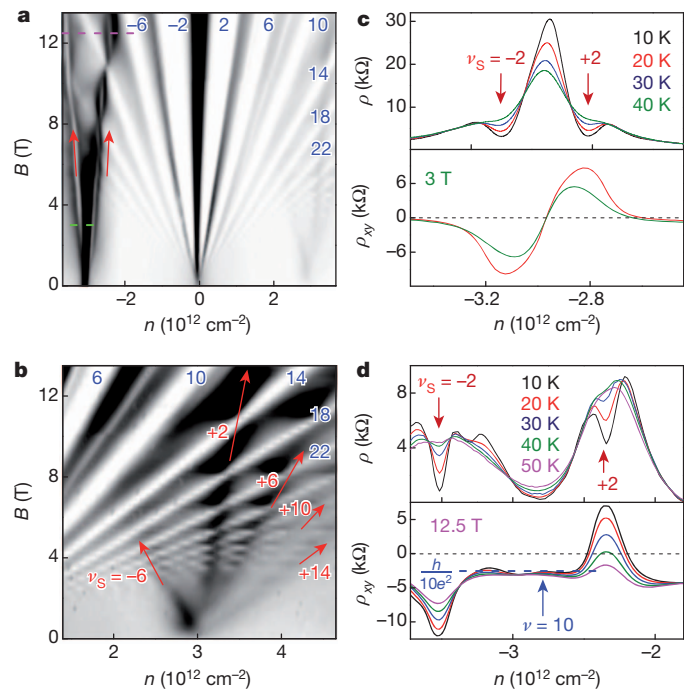
<sup>1</sup>School of Physics and Astronomy, University of Manchester, Manchester M13 9PL, UK. <sup>2</sup>Centre for Mesoscience and Nanotechnology, University of Manchester, Manchester M13 9PL, UK. <sup>3</sup>Physics Department, Lancaster University, Lancaster LA1 4YB, UK. <sup>4</sup>Laboratoire National des Champs Magnétiques Intenses, CNRS-UJF-UPS-INSa, F-38042 Grenoble, France. <sup>5</sup>Instituto de Ciencia de Materiales de Madrid, Sor Juana Inés de la Cruz 3, Madrid 28049, Spain.



**Figure 1 | Transport properties of Dirac fermions in moiré superlattices.**

**a**, Longitudinal resistivity,  $\rho_{xx}$ , as a function of  $n$ . Positive and negative values of  $n$  correspond to electrons and holes, respectively. The hole-side neutrality point shows a strong  $T$  dependence (Supplementary Information, section 1). Inset, one of the possible scenarios for the reconstruction of graphene's spectrum<sup>22</sup>. The band structure is plotted only for the first and second SBZ (shown in brown and green, respectively). Secondary Dirac cones appear in both conduction and valence bands at the edges of the SBZ, shown by the black hexagon. Where the cones merge, van Hove singularities appear in the density of states (for details, see ref. 22). **b**, The Hall resistivity,  $\rho_{xy}$ , changes sign at high electron and hole doping, revealing well-isolated secondary Dirac points. The data are for device A, for which  $n_S \approx 3.0 \times 10^{12} \text{ cm}^{-2}$ , yielding  $D \approx 12 \text{ nm}$ . We fabricated 11 aligned devices, six of which had essentially the same behaviour as shown here. The only difference was in  $n_S$ , which varied between  $3.0 \times 10^{12}$  and  $3.8 \times 10^{12} \text{ cm}^{-2}$  ( $D \approx 11 \pm 1 \text{ nm}$ ). One other device had  $n_S \approx 7.1 \times 10^{12} \text{ cm}^{-2}$ , which implies<sup>12</sup> that  $\theta \approx 1.2^\circ$  and required gate voltages close to dielectric breakdown. Inset, conductive atomic force microscope image of the moiré pattern for one of our devices. The centre-to-centre separation between the white spots is  $\sim 11 \text{ nm}$ . We note that, in measurements of  $\rho_{xy}$ , both positive and negative values of  $B$  were always used to symmetrize the data and subtract a small contribution due to  $\rho_{xx}$ .

to the superlattice filling factors  $\nu_S = \pm 2$ , where the carrier density is counted from the hSNP. In the middle of each maximum, there is a deep minimum (narrow white stripes in Fig. 2a). The minima in  $\rho_{xx}$  are accompanied by positive and negative extrema in  $\rho_{xy}$  (Fig. 2c, d). This shows that electron-like cyclotron trajectories in graphene's valence band persist when  $B$  is quantizing (that is, when it quantizes the spectrum). With decreasing  $T$ ,  $\rho_{xx}$  inside the narrow minima tends to zero and the corresponding extrema in  $\rho_{xy}$  become increasingly more pronounced, which is behaviour characteristic of the development of Shubnikov–de Haas oscillations into QHE states (Fig. 2c, d). The  $T$  dependence yields a QHE gap of  $\sim 20 \text{ meV}$  (Supplementary Fig. 6).

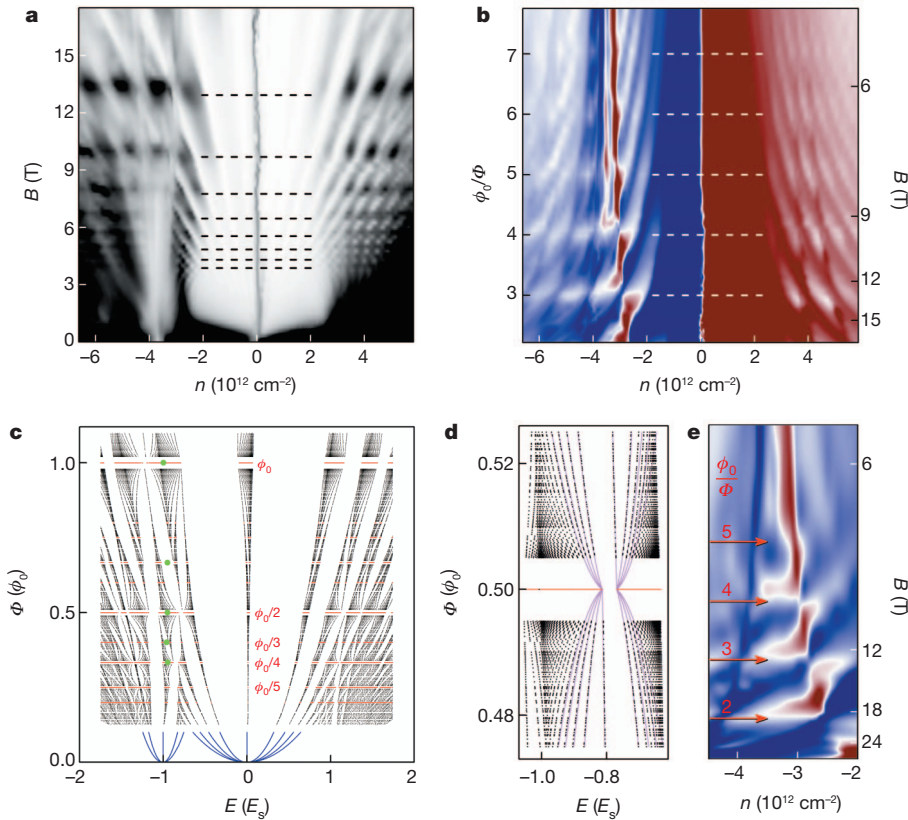


**Figure 2 | Quantization in graphene superlattices.** **a**, Longitudinal resistivity,  $\rho_{xx}(n, B)$ , at 20 K. Grey scale: white, 0 kΩ; black, 8.5 kΩ. **b**, Magnified view near the electron-side secondary neutrality point. Grey scale: white, 0 kΩ; black, 1.1 kΩ. Blue numbers denote  $\nu$  for the QHE states extending from the main Dirac point. The red arrows in **a** mark the superlattice quantum states that evolve along  $\nu_S = \pm 2$  (the arrows are shifted so as not to obscure the white stripes). In **b**, the red arrows indicate  $\nu_S$  for the electron-side neutrality point. It is difficult to associate the electron-side superlattice Landau levels decisively with any particular value of  $\nu_S$ , although the strongest peak in  $\rho_{xx}$  corresponds closely to  $\nu_S = +2$ . **c**, **d**, Detailed behaviour near the hSNP in fields marked by the dashed purple and green lines in **a**. Data are for device A but the same quantization behaviour was found in all devices that exhibited the secondary neutrality points. An exception is the white stripes at  $\nu_S = \pm 2$ , which were often smeared by inhomogeneity such that only broader maxima in  $\rho_{xx}$  remained (similar to the curves at 50 K). However, the narrow extrema in  $\rho_{xy}$ , associated with the minima in  $\rho_{xx}$  (**d**), were always present.  $h$ , Planck's constant;  $e$ , electron charge.

Unlike cyclotron gaps, this one is practically independent of  $B$ , as is seen also from the fact that the white stripes in Fig. 2a do not widen. With increasing  $T$ , the QHE states at  $\nu_S = \pm 2$  gradually disappear below 50 K but the maxima in  $\rho_{xx}$  persist up to 150 K.

Another notable feature of the observed fan diagrams are the multiple peaks in  $\rho_{xx}$  accompanied by zeros or deep minima in  $\rho_{xy}$ . This is seen most clearly for devices where doping sufficiently higher than  $n_S$  can be achieved (Fig. 3). Furthermore, in all our devices near the hSNP,  $\rho_{xy}$  repeatedly changes its sign with increasing  $B$ , indicating recurrent appearance and disappearance of electron-like orbits within graphene's valence band (Fig. 3b, e and Supplementary Figs 4 and 5). This means that, for a given  $n$ , the magnetic field alone can repeatedly generate new neutrality points. Such 'third-generation' neutrality points occur periodically as  $n$  and  $1/B$  vary, and form distinct groups characterized by particular values of  $1/B$  (Fig. 3 and Supplementary Information, section 3). Their periodicity in  $1/B$  is accurately described by unit fractions,  $\phi_0/q$ , of the magnetic flux,  $\Phi = BS_\otimes$ , per superlattice unit cell area,  $S_\otimes$ , where  $q$  is integer. In the conduction band, the fan diagrams also exhibit Landau levels extending from the secondary Dirac point, and numerous third-generation neutrality points with the same  $1/B$  periodicity are visible in Figs 2b and 3a, b. These features are weaker than those in the valence band. For example, the resistivity peak at the electron-side neutrality point is no longer there for  $B \approx 1 \text{ T}$  (Fig. 2b) and we did not observe the secondary QHE in the conduction band.

The observed superlattice behaviour suggests that complex spectral changes are induced by quantizing  $B$ . Theoretically, the problem is



**Figure 3 | Zak-type cloning of third-generation Dirac points.**

**a**, Longitudinal conductivity,  $\sigma_{xx} = \rho_{xx}/(\rho_{xx}^2 + \rho_{xy}^2)$ , as a function of  $n$  and  $B$  (device B). Grey scale: white, 0 k $\Omega$ ; black, 2.2 mS. The dashed lines indicate  $B_q = (1/q)\phi_0/S_\otimes$  with  $q = 3$ –10, where  $S_\otimes$  is determined from the measured  $n_S \approx 3.8 \times 10^{12} \text{ cm}^{-2}$  as  $S_\otimes = 4/n_S$ , which corresponds to the complete filling of the first SBZ<sup>22</sup>. **b**, Hall resistivity,  $\rho_{xy}$ , for the same device as a function of  $n$  and  $1/B$  (the latter expressed as  $\phi_0/\Phi$  on the left axis). The dashed white lines show the periodicity of Zak oscillations. Colour scale: navy,  $-2 \text{ k}\Omega$ ; white, 0 k $\Omega$ ; wine, 2 k $\Omega$ . In both plots,  $T = 2 \text{ K}$ . **c**, Hofstadter-like butterfly for the graphene-on-hBN superlattice. The electronic states are calculated following the approach in ref. 24 and are shown by black dots. For simple fractions  $p/q$ , we plot energies of the states in red. The regions around  $\Phi = \phi_0 p/q$  are empty

somewhat similar to that originally discussed by Zak<sup>23</sup> and Hofstadter<sup>9</sup> and later considered for two-dimensional electron systems<sup>3</sup> in semiconductor superlattices and for Dirac fermions in twisted bilayers<sup>24</sup>. The most general, but not proven, prediction is that superlattice spectra should be ‘self-similar’; that is, they should consist of multiple clones of an original spectrum, which appear at values of  $B$  such that  $\Phi = \phi_0(p/q)$ , where both  $p$  and  $q$  are integer. Our case of graphene on hBN is analysed in Supplementary Information, and the main theoretical results are summarized in Fig. 3c, d.

Figure 3c shows that the superlattice potential results in additional structure within each Landau level, which effectively broadens them with increasing  $B$ . The structured Landau levels extending from the main and secondary Dirac points strongly mix at high doping,  $|E| \geq E_S = \hbar v_F(\pi n_S)^{1/2}$ . The resulting pattern is different from that in semiconductor superlattices with a parabolic spectrum and weak modulation<sup>3,7,8</sup>. In the latter case, the fractal structure within each Landau level can be described by the original Hofstadter butterfly<sup>9</sup>, which appears periodically as a function of  $\phi_0/\Phi$ . In our case, in which there is a Dirac-like spectrum and strong modulation, the fractal pattern depends on the Landau level index,  $N$ , and  $B$  (Supplementary Information, section 5).

The calculated spectrum allows us to understand many features observed experimentally. Indeed, Fig. 3c shows a self-similarity such that magnetic states tend to entwine at  $\Phi = \phi_0 p/q$ , forming the fractal structure of the pattern. The strongest entwining occurs for unit fractions

because the corresponding supercells are too large for us to do the necessary calculations<sup>22,24</sup>. The blue curves show several low Landau levels for small  $B$ , which were calculated analytically for main and secondary Dirac fermions with parameters of the zero- $B$  spectrum. The green dots indicate the position of the Fermi level for  $n = -n_S$  (Supplementary Information). **d**, Section of **c** with superimposed Landau levels calculated as functions of  $\delta B$  (magenta). **e**, Hall resistivity,  $\rho_{xy}(n, 1/B)$ , for device C ( $n_S \approx 3.6 \times 10^{12} \text{ cm}^{-2}$ ) measured up to 29 T. The left and right axes are as in **b**. The same oscillatory behaviour is found for all our devices and is seen most clearly near the hSNP, where new neutrality points appear periodically for  $\phi_0/\Phi = q$ , as shown by the arrows. Colour scale: navy,  $-3 \text{ k}\Omega$ ; white, 0 k $\Omega$ ; wine, 3 k $\Omega$ .  $T = 20 \text{ K}$ . Full plots of  $\rho_{xy}(n, B)$  for devices B and C are provided in Supplementary Information, section 3.

(that is,  $p = 1$ ), and this results in an overall periodicity in  $1/B$  with a period of  $S_\otimes/\phi_0$ , in agreement with the experiment (Fig. 3a, b). The periodicity can be traced to the fact that for  $\Phi = \phi_0/q$  the system can be considered a new superlattice that has a unit cell  $q$  times larger than the original and that is placed in zero effective magnetic field<sup>23</sup>. An example of the resulting magneto-electronic (Zak) bands<sup>23</sup> is given in Supplementary Fig. 8.

We find that in our case Zak bands feature slightly gapped Dirac spectra. This finding is illustrated in Fig. 3d, which, as an example, magnifies a part of Fig. 3c near the hSNP and  $\Phi = \phi_0/2$ . Using the Zak spectrum calculated for  $\Phi = \phi_0/2$ , we can obtain its Landau quantization in small reduced fields,  $\delta B = B - B_2$ , where  $B_2 = \phi_0/2S_\otimes$  is the zero effective field acting on the cloned Dirac spectrum. The shape of the resulting Landau levels is given by  $\pm \sqrt{N|\delta B| + \mathcal{A}^2}$ , where  $\mathcal{A}$  is the gap in the local Dirac-like spectrum (Fig. 3d, magenta curves). Outside the empty horizontal region inaccessible in our numerical calculations (Supplementary Information, section 5), the Landau quantization of the Zak spectrum in  $\delta B$  yields practically the same electronic states as shown by the calculated points (Fig. 3d, black dots). Similar local Dirac spectra are found in other parts of the moiré butterfly<sup>24</sup> for all  $B_q = \phi_0/qS_\otimes$ , in agreement with the numerous third-generation neutrality points seen in the experiment. Using the calculated spectra, we have also determined the occupancy of Zak minibands for the case of four holes per moiré supercell (that is, at the hSNP) and found that for  $\Phi = \phi_0/q$  the Fermi energy lies inside the



corresponding Zak minibands, whereas for  $\Phi = \phi_0/(q + 1/2)$  it lies inside gaps. This explains the experimentally observed oscillations in  $\sigma_{xx}(B)$ .

Another experimental feature revealing Zak minibands and the hierarchy of superlattice gaps are the prominent QHE gaps at  $\nu_s = \pm 2$  near the hSNP (Fig. 2). For small  $B$ , they can be considered a result of Landau quantization for secondary Dirac fermions, and their zero Landau levels become separate from the rest of the spectrum (Fig. 3c). For higher  $B$ , the resulting gaps saturate, being limited in size by the presence of van Hove singularities at the edges of the SBZ (Fig. 1a). As Zak minibands become increasingly more pronounced the secondary Landau levels intertwine with main Landau levels and, at high doping, become indistinguishable from them (Fig. 3c). Therefore, the complex pattern of Landau levels in Fig. 3 at high doping can no longer be interpreted in terms of Landau quantization of either main or secondary Dirac fermions. The pattern becomes a 'Hofstadter–Landau' butterfly, specific to our strong-modulation regime and the linear spectrum. The largest fractal gaps near the hole-side Dirac point in Fig. 3c are in agreement with the  $\nu_s = \pm 2$  QHE states observed experimentally, which have activation energies almost independent of  $B$ . This behaviour is different from the case of weak modulation in semiconductor superlattices<sup>6–8</sup>, where Landau levels become structured but do not intertwine. However, increasing  $B$  such that  $\Phi > \phi_0$  can drive graphene superlattices into the regime of weak modulation (Supplementary Fig. 9). This regime is outside the range of  $B$  values available in our experiment. In addition to the large fractal gaps, our experimental data also reveal reproducible small-scale structure that cannot be traced back to either main or secondary neutrality points (see, for example, Fig. 3a, b near the hSNP at high  $B$  and Supplementary Fig. 5a). We attribute these fine features to further fractalization of the superlattice spectrum such that isolated Landau levels for the third-generation Dirac clones start being resolved (Fig. 3d). This is similar to the intra-Landau-level features reported in semiconductor superlattices<sup>7,8</sup> and warrants further investigation.

Graphene superlattices can be reliably fabricated for various types of transport measurement. This opens new lines of enquiry; in particular, the fractal quantization leads to such rich behaviour that its full understanding will require much further work, both theoretical and experimental. The demonstrated possibility of creating gaps at specifically chosen energies by controllably rotating graphene or other two-dimensional crystals within van der Waals heterostructures<sup>25</sup> can be used to design novel electronic and optoelectronic devices.

## METHODS SUMMARY

Our devices were multiterminal Hall bars fabricated following the procedure described in ref. 27. In brief, monolayer graphene was deposited on top of a relatively thick ( $>30$  nm) hBN crystal<sup>28</sup> and then covered with another hBN crystal. The encapsulation protects the graphene from the environment and allows high  $\mu$  values, little residual doping ( $<10^{11}$  cm<sup>-2</sup>) and little charge inhomogeneity<sup>27</sup>. The interfaces between the graphene and the hBN were atomically clean over the entire active device area<sup>25</sup>. The whole stack was assembled on top of an oxidized Si wafer that served as a back gate. To align the crystal lattices, we used an optical microscope to choose straight edges of graphene and hBN crystallites, which indicate the principal crystallographic directions (see, for example, figure 2 of ref. 30). During the assembly, the graphene was rotated relatively to the bottom hBN crystal to make their edges parallel. We estimate our alignment precision to be  $\sim 1^\circ$ . The top hBN crystal was then rotated by  $\sim 15^\circ$  with respect to the aligned edges, which ensured no spectral reconstruction at  $E_s < 1$  eV due to the second graphene–hBN interface.

In practice, gate dielectric breakdown for oxidized silicon wafers occurs in fields less than  $0.4$  V nm<sup>-1</sup>, and this limits achievable values of  $n$  to less than  $7 \times 10^{12}$  cm<sup>-2</sup> ( $E_s < 0.35$  eV). Accordingly, the observation of secondary Dirac points requires alignment with  $\theta \leq 1^\circ$  (ref. 12). For random deposition of graphene on hBN, the probability of finding transport devices with superlattice features can be estimated to be only a few per cent, even if high gate voltages are used, which is rare given the desire to avoid accidental breakdown. Previously, we investigated more than 25 graphene-on-hBN devices<sup>27</sup> and none of them exhibited any sign of superlattice effects. This shows that careful alignment is essential for the observation of secondary Dirac spectra in transport measurements.

Received 4 December 2012; accepted 3 April 2013.

Published online 15 May 2013.

- Weiss, D., Klitzing, K. V., Ploog, K. & Weimann, G. Magnetoresistance oscillations in a two-dimensional electron gas induced by a submicrometer periodic potential. *Europhys. Lett.* **8**, 179–184 (1989).
- Weiss, D. *et al.* Electron pinball and commensurate orbits in a periodic array of scatterers. *Phys. Rev. Lett.* **66**, 2790–2793 (1991).
- Pfannkuche, D. & Gerhardt, R. R. Theory of magnetotransport in two-dimensional electron systems subjected to weak two-dimensional superlattice potentials. *Phys. Rev. B* **46**, 12606–12626 (1992).
- Ferry, D. K. Quantum magnetotransport in lateral surface superlattices. *Prog. Quantum Electron.* **16**, 251–317 (1992).
- Albrecht, C. *et al.* Fermiology of two-dimensional lateral superlattices. *Phys. Rev. Lett.* **83**, 2234–2237 (1999).
- Schlösser, T., Ensslin, K., Kotthaus, J. P. & Holland, M. Internal structure of a Landau band induced by a lateral superlattice: a glimpse of Hofstadter's butterfly. *Europhys. Lett.* **33**, 683–688 (1996).
- Albrecht, C. *et al.* Evidence of Hofstadter's fractal energy spectrum in the quantized Hall conductance. *Phys. Rev. Lett.* **86**, 147–150 (2001).
- Geisler, M. C. *et al.* Detection of a Landau band-coupling-induced rearrangement of the Hofstadter butterfly. *Phys. Rev. Lett.* **92**, 256801 (2004).
- Hofstadter, D. R. Energy levels and wave functions of Bloch electrons in rational and irrational magnetic fields. *Phys. Rev. B* **14**, 2239–2249 (1976).
- Xue, J. M. *et al.* Scanning tunnelling microscopy and spectroscopy of ultra-flat graphene on hexagonal boron nitride. *Nature Mater.* **10**, 282–285 (2011).
- Decker, R. *et al.* Local electronic properties of graphene on a BN substrate via scanning tunneling microscopy. *Nano Lett.* **11**, 2291–2295 (2011).
- Yankowitz, M. *et al.* Emergence of superlattice Dirac points in graphene on hexagonal boron nitride. *Nature Phys.* **8**, 382–386 (2012).
- Park, C. H., Yang, L., Son, Y. W., Cohen, M. L. & Louie, S. G. New generation of massless Dirac fermions in graphene under external periodic potentials. *Phys. Rev. Lett.* **101**, 126804 (2008).
- Bliokh, Y. P., Freilikher, V., Savel'ev, S. & Nori, F. Transport and localization in periodic and disordered graphene superlattices. *Phys. Rev. B* **79**, 075123 (2009).
- Tiwari, R. P. & Stroud, D. Tunable band gap in graphene with a noncentrosymmetric superlattice potential. *Phys. Rev. B* **79**, 205435 (2009).
- Barbier, M., Vasilopoulos, P. & Peeters, F. M. Extra Dirac points in the energy spectrum for superlattices on single-layer graphene. *Phys. Rev. B* **81**, 075438 (2010).
- Burset, P., Yeyati, A. L., Brey, L. & Fertig, H. A. Transport in superlattices on single-layer graphene. *Phys. Rev. B* **83**, 195434 (2011).
- Wu, S., Kili, M. & Paramakanti, A. Graphene under spatially varying external potentials: Landau levels, magnetotransport, and topological modes. *Phys. Rev. B* **85**, 195404 (2012).
- Ortiz, C., Yang, L. & van den Brink, J. Graphene on incommensurate substrates: trigonal warping and emerging Dirac cone replicas with halved group velocity. *Phys. Rev. B* **86**, 081405 (2012).
- Zarenia, M., Leenaerts, O., Partoens, B. & Peeters, F. M. Substrate-induced chiral states in graphene. *Phys. Rev. B* **86**, 085451 (2012).
- Kindermann, M. M., Uchoa, B. & Miller, D. L. Zero energy modes and gate-tunable gap in graphene on hexagonal boron nitride. *Phys. Rev. B* **86**, 115415 (2012).
- Wallbank, J. R., Patel, A. A., Mucha-Kruczynski, M., Geim, A. K. & Fal'ko, V. I. Generic miniband structure of graphene on a hexagonal substrate. Preprint at <http://arxiv.org/abs/1211.4711> (2012).
- Zak, J. Magnetic translation group. *Phys. Rev.* **134**, A1602–A1606 (1964).
- Bistritzer, R. & MacDonald, A. H. Moire butterflies. *Phys. Rev. B* **84**, 035440 (2011).
- Haigh, S. J. *et al.* Cross-sectional imaging of individual layers and buried interfaces of graphene-based heterostructures and superlattices. *Nature Mater.* **11**, 764–767 (2012).
- Li, G., Luican, A. & Andrei, E. Y. Scanning tunneling spectroscopy of graphene on graphite. *Phys. Rev. Lett.* **102**, 176804 (2009).
- Mayorov, A. S. *et al.* Micrometer-scale ballistic transport in encapsulated graphene at room temperature. *Nano Lett.* **11**, 2396–2399 (2011).
- Dean, C. R. *et al.* Boron nitride substrates for high-quality graphene electronics. *Nature Nanotechnol.* **5**, 722–726 (2010).
- Mayorov, A. S. *et al.* How close can one approach the Dirac point in graphene experimentally? *Nano Lett.* **12**, 4629–4634 (2012).
- Geim, A. K. & Novoselov, K. S. The rise of graphene. *Nature Mater.* **6**, 183–191 (2007).

Supplementary Information is available in the online version of the paper.

**Acknowledgements** We thank D. Weiss, A. MacDonald and F. Peeters for discussions. This work was supported by the European Research Council, the Körber Foundation, the Office of Naval Research, the Air Force Office of Scientific Research and the Royal Society.

**Author Contributions** R.V.G., L.A.P. and A.K.G. designed the project. R.V.G. and R.J. made the graphene devices. G.L.Y., D.C.E., L.A.P. and A.S.M. carried out the measurements. K.S.N., A.M., C.R.W., B.A.P., M.P. and I.V.G. provided experimental support. V.I.F., A.A.P., J.R.W., M.M.-K., A.K.G. and F.G. developed the theory. A.K.G. and V.I.F. wrote the manuscript with input from I.V.G., R.V.G., L.A.P., K.S.N. and F.G. All authors participated in discussions.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.V.G. ([blizza@gmail.com](mailto:blizza@gmail.com)).

# Hofstadter's butterfly and the fractal quantum Hall effect in moiré superlattices

C. R. Dean<sup>1</sup>, L. Wang<sup>2</sup>, P. Maher<sup>3</sup>, C. Forsythe<sup>3</sup>, F. Ghahari<sup>3</sup>, Y. Gao<sup>2</sup>, J. Katoch<sup>4</sup>, M. Ishigami<sup>4</sup>, P. Moon<sup>5</sup>, M. Koshino<sup>5</sup>, T. Taniguchi<sup>6</sup>, K. Watanabe<sup>6</sup>, K. L. Shepard<sup>7</sup>, J. Hone<sup>2</sup> & P. Kim<sup>3</sup>

Electrons moving through a spatially periodic lattice potential develop a quantized energy spectrum consisting of discrete Bloch bands. In two dimensions, electrons moving through a magnetic field also develop a quantized energy spectrum, consisting of highly degenerate Landau energy levels. When subject to both a magnetic field and a periodic electrostatic potential, two-dimensional systems of electrons exhibit a self-similar recursive energy spectrum<sup>1</sup>. Known as Hofstadter's butterfly, this complex spectrum results from an interplay between the characteristic lengths associated with the two quantizing fields<sup>1–10</sup>, and is one of the first quantum fractals discovered in physics. In the decades since its prediction, experimental attempts to study this effect have been limited by difficulties in reconciling the two length scales. Typical atomic lattices (with periodicities of less than one nanometre) require unfeasibly large magnetic fields to reach the commensurability condition, and in artificially engineered structures (with periodicities greater than about 100 nanometres) the corresponding fields are too small to overcome disorder completely<sup>11–17</sup>. Here we demonstrate that moiré superlattices arising in bilayer graphene coupled to hexagonal boron nitride provide a periodic modulation with ideal length scales of the order of ten nanometres, enabling unprecedented experimental access to the fractal spectrum. We confirm that quantum Hall features associated with the fractal gaps are described by two integer topological quantum numbers, and report evidence of their recursive structure. Observation of a Hofstadter spectrum in bilayer graphene means that it is possible to investigate emergent behaviour within a fractal energy landscape in a system with tunable internal degrees of freedom.

The total number of electron states per area of a completely filled Bloch band is  $n_0 = 1/A$ , where  $A$  is the area of the unit cell of the periodic potential. In a magnetic field,  $B$ , the number of states per area of each filled Landau level is given by  $B/\phi_0$ , where  $\phi_0 = h/e$  is the magnetic flux quantum ( $h$ , Planck's constant;  $e$ , magnitude of the electron charge). The quantum description of electrons subjected simultaneously to both a periodic electric field and a magnetic field can be simply parameterized by the dimensionless ratio  $\phi/\phi_0$ , where  $\phi = BA$  is the magnetic flux per unit cell. The general solution, however, exhibits a rich complexity due to the incommensurate periodicities of the Bloch and Landau states<sup>18</sup>. For commensurate fields, corresponding to rational values of  $\phi/\phi_0 = p/q$ , where  $p$  and  $q$  are co-prime integers, the single-particle Bloch band splits into  $q$  subbands<sup>1</sup> (beginning with the Landau level description, it can be shown that, equivalently, the energy diagram is parameterized by  $\phi_0/\phi = q/p$  such that at these same rational values each Landau level splits into  $p$  subbands<sup>2</sup>). This results in a quasi-continuous distribution of incommensurate quantum states with a self-similar recursive structure, yielding a butterfly-like fractal energy diagram called the Hofstadter butterfly<sup>1</sup> (Supplementary Information).

Important insight into this system came from consideration of the density of charge carriers,  $n$ , required to fill each fractal subband<sup>2</sup>.

Replotting the Hofstadter energy spectrum as integrated density versus field shows that all spectral gaps are constrained to linear trajectories in the density–field diagram (Wannier diagram). This can be described by a simple Diophantine relation

$$(n/n_0) = t(\phi/\phi_0) + s \quad (1)$$

where  $n/n_0$  and  $\phi/\phi_0$  are the normalized carrier density and magnetic flux, respectively, and  $s$  and  $t$  are both integer valued. Here  $n/n_0$  represents the Bloch band filling fraction, which is distinct from the usual Landau level filling fraction,  $\nu = n\phi_0/B$  (the two are related by the normalized flux, that is,  $n/n_0 = \nu\phi/\phi_0$ ). The physical significance of the quantum numbers  $s$  and  $t$  became fully apparent with the discovery of the integer quantum Hall effect<sup>19</sup> (QHE), after which it was shown that the Hall conductivity associated with each minigap in the fractal spectrum is quantized according to the relation  $\sigma_{xy} = te^2/h$  (refs 3, 4). The second quantum number,  $s$ , physically corresponds to the Bloch band filling index in the fractal spectrum<sup>5</sup>. This formalism suggests several unique and unambiguous experimental signatures associated with the Hofstadter energy spectrum that are distinct from the conventional QHE. First, the Hall conductance can vary non-monotonically and can even fluctuate in sign. Second, the Hall conductance plateaux, together with vanishing longitudinal resistance, can appear at non-integer Landau level filling fractions. Third, the Hall conductance plateaux remain quantized in integral multiples of  $e^2/h$ . However, the quantization integer is not directly associated with the usual Landau level filling fraction. Instead, quantization is equal to the slope of the gap trajectory in the  $n/n_0$ – $\phi/\phi_0$  Wannier diagram, in accordance with the Diophantine equation (equation (1)).

Minigaps within the fractal energy spectrum become significant only once the magnetic length ( $l_B = \sqrt{\hbar/eB}$ ;  $\hbar$ , Planck's constant divided by  $2\pi$ ), which characterizes the cyclotron motion, is of the same order as the wavelength of the periodic potential, which characterizes the Bloch waves. For usual crystal lattices, where the interatomic spacing is a few ångströms, the necessary magnetic field is unfeasibly large, in excess of 10,000 T. The main experimental effort therefore has been to lithographically define artificial superlattices<sup>11–17</sup> with unit-cell dimensions of order tens of nanometres, so that the critical magnetic field remains small enough to be achievable in the lab yet still large enough that the QHE is fully resolved without being smeared out by disorder. Fabricating the optimally sized periodic lattice while maintaining coherent registry over the full device and without introducing substantial disorder has proven to be a formidable technical challenge. Patterned GaAs/AlGaAs heterostructures with  $\sim 100$ -nm-period gates provided the first experimental support for the predictions of a Hofstadter spectrum<sup>14–16</sup>. However, limited ability to tune the carrier density or reach the fully developed QHE regime in these samples has made it difficult to map out the complete spectrum. Although similar concepts have been pursued in non-solid-state model systems<sup>20,21</sup>, the rich physics of the Hofstadter spectrum remains largely unexplored.

<sup>1</sup>Department of Physics, The City College of New York, New York, New York 10031, USA. <sup>2</sup>Department of Mechanical Engineering, Columbia University, New York, New York 10027, USA. <sup>3</sup>Department of Physics, Columbia University, New York, New York 10027, USA. <sup>4</sup>Department of Physics and Nanoscience Technology Center, University of Central Florida, Orlando, Florida 32816-2385, USA. <sup>5</sup>Department of Physics, Tohoku University, Sendai 980-8578, Japan. <sup>6</sup>National Institute for Materials Science, 1-1 Namiki, Tsukuba 305-0044, Japan. <sup>7</sup>Department of Electrical Engineering, Columbia University, New York, New York 10027, USA.

Heterostructures consisting of atomically thin materials in a multi-layer stack provide a new means of realizing a two-dimensional system with a laterally modulated periodic structure. In particular, coupling between graphene and hexagonal boron nitride (hBN), whose crystal lattices are isomorphic, results in a periodic moiré pattern. The moiré wavelength is directly related to the angular rotation between the two lattices<sup>22–24</sup>, and can be tuned through the desired length scales without the need for lithographic techniques<sup>8,9</sup>. Moreover, hBN provides an ideal substrate for achieving high-mobility graphene devices, which is crucial for high-resolution quantum Hall measurements<sup>25,26</sup>, and field-effect gating in graphene allows the Fermi energy to be continuously varied through the entire moiré Bloch band.

In this study, we used Bernal-stacked bilayer graphene (BLG) Hall bars fabricated on hBN substrates (Fig. 1a, b) using mechanical exfoliation followed by co-lamination (Methods Summary). Figure 1b shows a non-contact atomic force microscopy (AFM) image acquired from an example device. In the magnified region, a triangular moiré pattern is visible with wavelength  $15.5 \pm 0.9$  nm. This is comparable to the maximal moiré wavelength of  $\sim 14$  nm expected for graphene on hBN<sup>22–24</sup>, suggesting that in this device the BLG lattice is oriented relative to the underlying hBN lattice with near-zero angle mismatch.

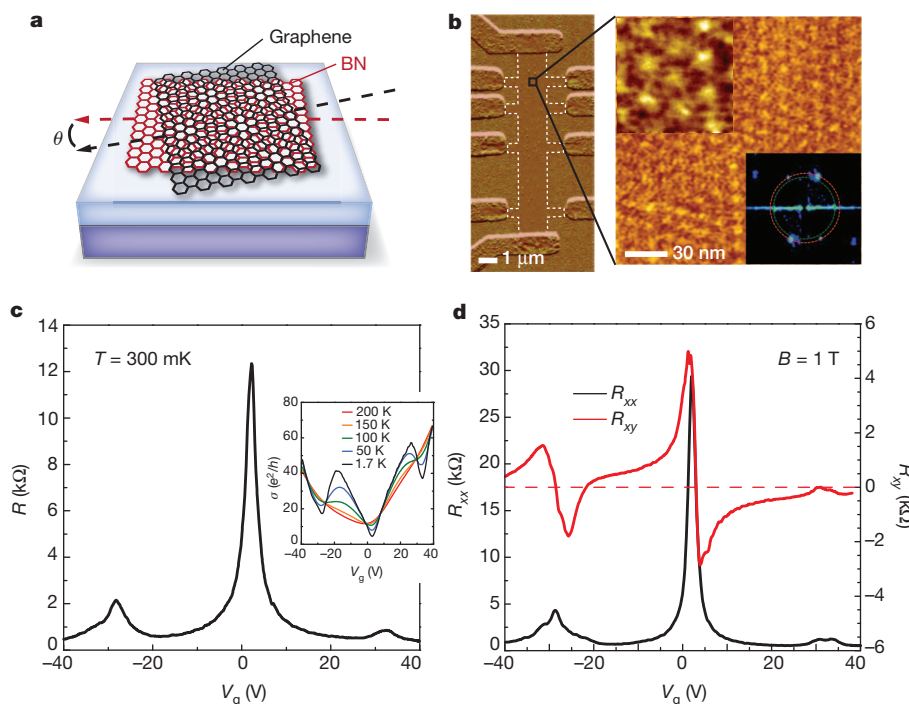
Figure 1c shows transport data measured from the same device. In addition to the usual resistance peak at the charge neutrality point (CNP), occurring at gate voltage  $V_g \approx 2$  V, two additional satellite resistance peaks appear, symmetrically located at  $V_{\text{satl}} \approx \pm 30$  V relative to the CNP. These satellite features are consistent with a depression in the density of states at the superlattice Brillouin zone band edge, analogous to previous spectroscopic measurements of single-layer graphene coupled to a moiré superlattice<sup>24,27</sup>. Assuming non-overlapping bands,  $|V_{\text{satl}}|$  gives an estimate of the moiré wavelength of  $\sim 14.6$  nm (Supplementary Information), in good agreement with the AFM measurements. The nature of these satellite peaks can be further probed in the semiclassical, low- $B$  transport regime. In Fig. 1d, longitudinal resistance,  $R_{xx}$ , and transverse Hall resistance,  $R_{xy}$ , are plotted versus gate voltage at

$B = 1$  T. Near the central CNP, the Hall resistance changes sign as the Fermi energy passes from the electron to the hole band. The same trend also appears near  $V_{\text{satl}}$ , consistent with the Fermi energy passing through a second band edge. This provides further confirmation that the moiré pattern, acting as a periodic potential superlattice, gives rise to a mini-Brillouin zone band<sup>28</sup>. We observed the satellite peak to be more developed in the hole branch than in the electron branch in all samples, in agreement with previous experimental and theoretical studies of hBN-supported monolayer graphene<sup>24,27,28</sup>. The satellite peaks vanish at temperatures above 100 K (Fig. 1c, inset), indicating that the coupling between the BLG and hBN atomic lattices is of order  $\sim 10$  meV. Perfect crystallographic alignment between graphene and hBN is expected to open a  $\sim 50$ -meV bandgap<sup>29,30</sup>, leading to a low-temperature divergence in the resistance at the CNP. The weak temperature dependence observed in our device suggests the absence of a gap, possibly owing to the lattice mismatch between the BLG and hBN.

In the remainder of this Letter, we focus on magnetotransport measured at high field. Figure 2a shows the evolution of  $R_{xx}$  and  $R_{xy}$  for magnetic fields up to 31 T. In the left panel (a Landau fan diagram),  $R_{xx}$  is plotted against the experimentally tunable gate voltage and magnetic field. In the right panel, the magnitude of the corresponding  $R_{xy}$  is plotted against the dimensionless parameters appearing in the Diophantine equation,  $n/n_0$  and  $\phi/\phi_0$ . This Wannier diagram is simply the Landau fan diagram with both axes relabelled by dimensionless units defined by normalizing to the moiré unit-cell area.

In a conventional quantum Hall system, the Landau fan diagram exhibits straight lines, tracking minima in  $R_{xx}$  and plateaux in  $R_{xy}$ . Plotted against  $n/n_0$  and  $\phi/\phi_0$ , the slope of each line is precisely the Landau level filling fraction,  $\nu$ , and all lines converge to the origin. White lines in Fig. 2a identify QHE states matching this description, tracking Landau level filling fractions  $\nu = 4, 8$  and  $12$ . This is consistent with the usual QHE hierarchy associated with a conventional degenerate BLG spectrum.

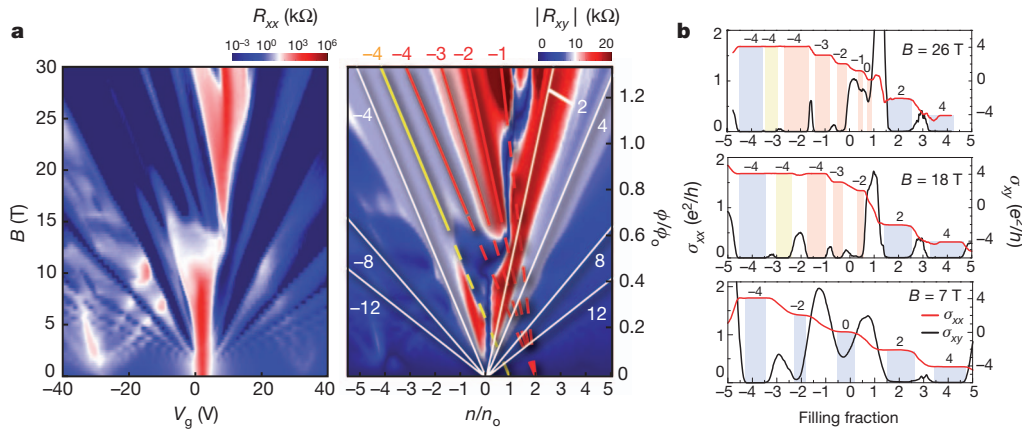
At large magnetic fields, several additional QHE states, exhibiting minima in  $R_{xx}$  together with plateaux in  $R_{xy}$ , develop outside the usual



**Figure 1 | Moiré superlattice.** **a**, Sketch of graphene on hBN showing the emergence of a moiré pattern. The moiré wavelength varies with the mismatch angle,  $\theta$ . **b**, Left: an AFM image of a multiterminal Hall bar. Right: a high-resolution image of a magnified region. The moiré pattern is evident as a triangular lattice (upper inset shows a further magnified region). A fast Fourier transform of the scan area (lower inset) confirms a triangular lattice symmetry

with periodicity  $15.5 \pm 0.9$  nm. **c**, Measured resistance versus gate voltage at zero magnetic field. Inset: the corresponding conductivity versus temperature, indicating that the satellite features disappear at temperatures greater than  $\sim 100$  K. **d**, Longitudinal resistance (left axis) and Hall resistance (right axis) versus gate voltage at  $B = 1$  T. The Hall resistance changes sign and passes through zero at the same gate voltage as the satellite peaks.





**Figure 2 | Emergence of anomalous quantum Hall states.** **a**, Landau fan diagram showing longitudinal resistance,  $R_{xx}$  (left), and Hall resistance,  $R_{xy}$  (right).  $R_{xx}$  is plotted versus magnetic field on the vertical axis and versus gate bias on the horizontal axis. In the diagram showing  $R_{xy}$ , the axes are scaled by the size of the moiré unit cell to give  $\phi/\phi_0$  on the vertical axis and  $n/n_0$  on the horizontal axis. QHE states corresponding to the conventional BLG spectrum are indicated by white lines. Solid yellow and red lines track the QHE outside this conventional spectrum, with dashed lines indicating the projected  $n/n_0$  intercept. The slope of each line is shown on the top axis. **b**, Longitudinal and

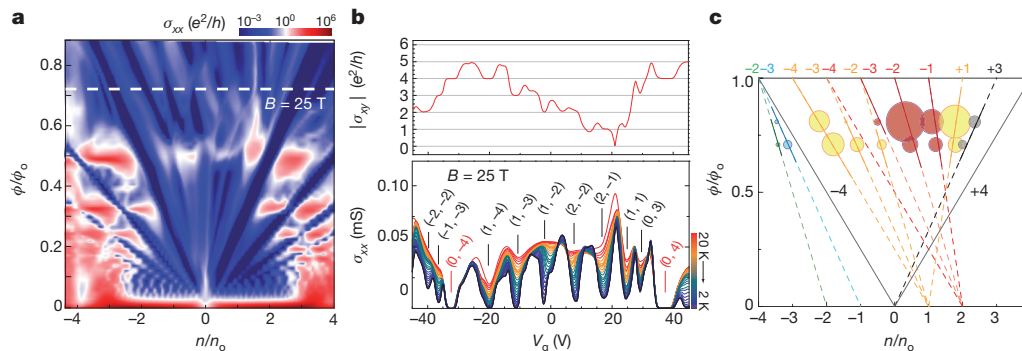
transverse Hall conductivities corresponding to line cuts at constant magnetic field (constant  $\phi/\phi_0$ ) from the Landau fan diagram in **a**. At  $B = 7$  T, the QHE ladder is consistent with previous reports for bilayer graphene. At  $B = 18$  and 26 T, additional QHE states emerge, showing Hall conductivity plateaux quantized in integer multiples of  $e^2/h$ , but appearing at non-integer Landau level filling fractions. Yellow and red bars indicate correspondence to the similarly coloured anomalous features marked by solid lines in **a**. Blue bars indicate the conventional QHE features. Numbers label the quantization integer for each plateau.

BLG sequence and also follow straight lines in the Landau fan diagram, but converge to non-zero values of  $n/n_0$ . Yellow and red lines in Fig. 2a trace examples of these anomalous QHE states appearing within the lowest Landau level. Unlike the conventional QHE states, each of the anomalous QHE states is characterized by both an integer-valued intercept,  $s$  (yellow and red lines converge to  $n/n_0 = 1$  and 2, respectively) and an integer-valued slope,  $t$  (labelled along the top axis in the figure). In Fig. 2b, longitudinal and Hall conductivities measured at constant magnetic field (corresponding to horizontal line cuts through the fan diagram in Fig. 2a) are plotted against Landau level filling fraction,  $\nu$ . At large magnetic fields, the anomalous QHE states are remarkably well developed, exhibiting wide plateaux in  $\sigma_{xy}$  concomitant with  $\sigma_{xx} = 0$ . Moreover, these states appear in general at non-integer filling fractions. Comparison between Fig. 2a and Fig. 2b further reveals that Hall conductivity plateaux are quantized in integer multiples of  $e^2/h$ , where the quantization integer  $t$  equals the slope in the Wannier diagram. Similar internal structure is observed within higher-order Landau levels (Fig. 3 and Supplementary Information). The anomalous QHE states observed here are consistent with fully developed

spectral gaps resulting from a Hofstadter-type energy spectrum. Moreover, our ability to map fully the density-field space provides a remarkable confirmation of the Diophantine equation: we observe direct evidence that QHE features associated with the Hofstadter spectral gaps are characterized by the two quantum numbers,  $s$  and  $t$ , corresponding to the  $n/n_0$  intercept and the slope, respectively, in the Wannier diagram.

Figure 3 shows similar data to Fig. 2, but measured from a separate device in which the moiré wavelength is only 11.6 nm. Again, QHE states appear outside the conventional Bernal-stacked BLG sequence and follow straight lines whose origin and slope are both integer valued, with the slope exactly matching the Hall quantization, in precise agreement with the Diophantine equation. Similar to the previous device, the  $\nu = 0$  insulating state undergoes a drastic change near  $\phi/\phi_0 = 1/2$ , when anomalous QHE states associated with the fractal gaps begin to develop fully.

In Fig. 3b, the lower panel shows  $\sigma_{xx}$  measured at  $B = 25$  T, corresponding to a horizontal line cut through Fig. 3a (dashed white line), for a variety of sample temperatures. The magnitudes of the fractal gaps



**Figure 3 | Fractal gaps.** **a**, Landau fan diagrams similar to those in Fig. 2 but measured from a separate device. Here the zero-field satellite peak position indicates a moiré period of 11.6 nm, indicating that the superlattice unit cell was approximately 1.5 times smaller in this device than in the one used in Fig. 2. Significantly more structure is observed here than in Fig. 2. **b**, Bottom: the evolution of  $\sigma_{xx}$  with temperature varying between 2 and 20 K, acquired at constant  $B = 25$  T, which corresponds to the line cut shown in **a**. Top: the corresponding  $\sigma_{xy}$  at  $T = 2$  K. The bracketed numbers label the  $(s, t)$  values of

the corresponding fractal gaps according to the Diophantine equation. **c**, Bubble plot of energy gaps determined from the temperature dependence calculated at two magnetic fields ( $B = 25$  and 28.5 T). The gaps are plotted as circles with radius scaled relative to the largest gap value measured. Dashed lines trace select fractal gap positions allowed by the Diophantine equation. Solid lines trace regions where the corresponding fractal gaps appear as minima in  $\sigma_{xx}$  together with quantized plateaux in  $\sigma_{xy}$ . The colours indicate gaps with the same quantum number  $s$ .

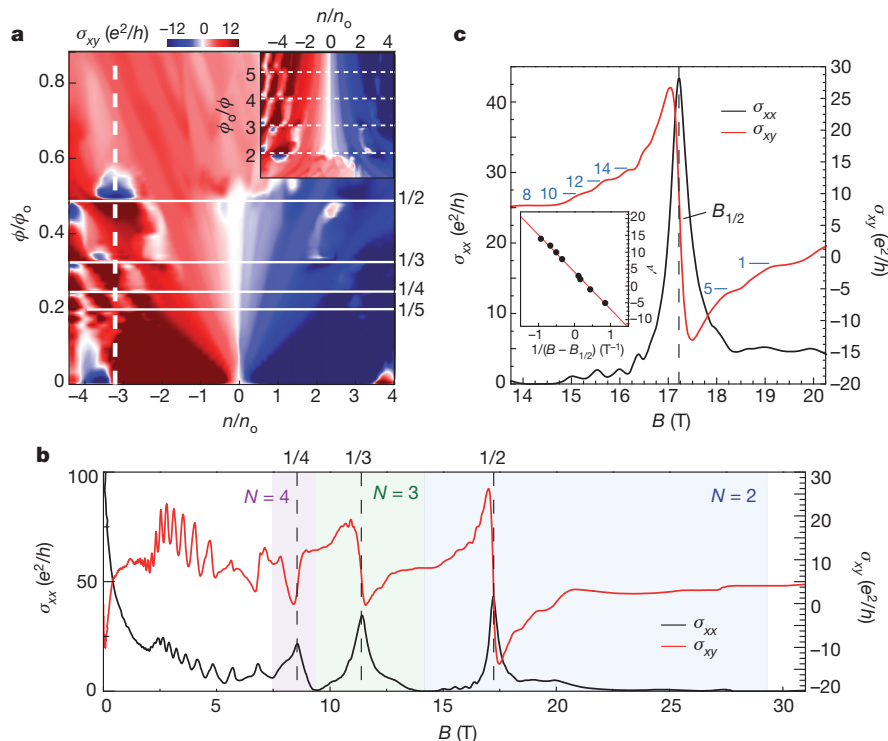
were estimated from the temperature dependence of the  $\sigma_{xx}$  minima in the thermally activated regime (Supplementary Information), at two separate magnetic field values, 25 and 28.5 T. Figure 3c summarizes our findings. Each fractal gap is marked by a circle centred at the corresponding  $(n/n_o, \phi/\phi_o)$  coordinates, and with radius scaled relative to the largest gap value (Supplementary Information). As the magnetic field increases, the spectral energies develop in a complicated way: some gaps grow with field (for example those with  $(s, t) = (1, 1)$  and  $(2, -2)$ ), whereas others shrink (for example  $(1, -4)$ ). For a fixed magnetic field, it seems generally true that for constant  $s$  values, fractal gap states exhibit increasing gap size as  $t$  increases. For example, at  $B = 25$  T,  $\Delta_{(1,-4)} \approx 48$  K whereas  $\Delta_{(1,-3)} \approx 30$  K. This contradicts the prediction that fractal gaps corresponding to lower quantum numbers have larger gap values<sup>1</sup>. We note that such a trend was subsequently found to be specific to square lattice symmetries<sup>2,5</sup>. Furthermore, a non-trivial case also arises when two fractal gap states overlap<sup>2</sup>, such as occurs between the  $(1, -2)$  and  $(2, -3)$  states in our data as  $\phi/\phi_o \rightarrow 1$ . Further theoretical analysis specific to moiré-patterned BLG is necessary to understand fully the trends highlighted here.

Figure 4a shows a normalized Landau fan diagram of  $R_{xy}$  values corresponding to the  $R_{xx}$  data in Fig. 3a. Dashed horizontal lines in the figure label special values of the normalized magnetic flux,  $\phi/\phi_o = 1/m$ , where  $m$  is integer valued. Referred to as the 'pure cases', these lines of high symmetry provide the framework for the recursive structure of the butterfly spectrum, marking the boundaries of the repeating sub-cells that appear within the main cell<sup>5</sup>. In Fig. 4a, at the pure cases in the fan diagram,  $R_{xy}$  seems to tend to zero and change sign. This is also seen from the single line trace in Fig. 4b. In the quantum Hall regime, the longitudinal conductivity has a local peak as the magnetic field passes through the pure cases, with the corresponding Hall conductivity exhibiting a sharp transition. For large magnetic fields, both of these

features span the full Landau level along lines of constant  $\phi/\phi_o$ , as seen in Figs 3a and 4a, respectively. Near the field corresponding to  $\phi/\phi_o = 1/2$ , labelled in Fig. 4c as  $B_{1/2}$ , plateaux appear in  $R_{xy}$  together with minima in  $R_{xx}$ , resembling a mini QHE series centred on  $B_{1/2}$ . If we redefine the local effective magnetic field as  $B' = B - B_{1/2}$ , then, according to the usual QHE formalism, we expect to find that  $\nu' = (1/B')n'h/e$ , where  $\nu'$  is an effective filling fraction given by the Hall quantization,  $B'$  is the value of the effective magnetic field at the  $R_{xx}$  minima and  $n'$  is an effective carrier density. The inset in Fig. 4c shows a plot of  $\nu'$  versus  $1/B'$  and the data indeed follows a linear trend. In spite of the large magnetic field ( $B_{1/2} \approx 17.3$  T), this indicates that locally the electrons behave as if the magnetic field is reduced to zero. We regard this as compelling evidence of the long-predicted recursive nature of the Hofstadter spectrum, where repeated mini fan diagrams emerge within the main one. We note that the linear trend shown inset in Fig. 4c does not pass through the origin, but is vertically offset by  $4.1 \pm 0.1$ . The origin of this offset is unclear but may be related to disorder effects because in this regime the spectrum is not fully gapped<sup>6</sup>.

Finally, we note that precise modelling of the fractal spectrum (Supplementary Information) requires a quantitative understanding of the atomic-scale couplings between BLG and hBN and between the BLG layers, which are not well known. Additionally the odd-integer quantum numbers observed in our experiment, which may not be accounted for in a purely single-particle picture, invite further investigation of correlated electron behaviour within the Hofstadter spectrum.

In conclusion, we report experimental confirmation that the generalized behaviour of electrons subjected simultaneously to both a magnetic field and a spatially varying periodic electrostatic field is described by a Diophantine equation involving two topological quantum numbers. The ability to tune the competing length scales governing these Landau–Bloch–Dirac states in graphene coupled to a substrate-induced



**Figure 4 | Recursive structure.** **a**,  $R_{xy}$  Wannier diagram for the device used in Fig. 3. White solid lines label  $\phi/\phi_o$  values corresponding to the pure cases,  $\phi/\phi_o = 1/m$ . Inset: data replotted against  $\phi/\phi_o$ , illustrating that the main experimental features exhibit a  $1/B$  periodicity. **b**, Longitudinal conductivity,  $\sigma_{xx}$  (left axis), and Hall conductivity,  $\sigma_{xy}$  (right axis), versus magnetic field, measured at a constant gate voltage ( $V_g = -39$  V, corresponding to the white dashed line in **a**). Blue, green and purple bands mark the boundaries of the

conventional  $N = 2, 3$  and 4 Landau levels, respectively. Dashed lines mark the values of  $B$  corresponding to the pure fractions labelled in **a**. The values of  $\phi/\phi_o$  are shown on the top axis. **c**, Details of both  $\sigma_{xx}$  and  $\sigma_{xy}$  in the vicinity of  $\phi/\phi_o = 1/2$  (**b**). Plateaux in  $\sigma_{xy}$  concomitant with minima in  $\sigma_{xx}$  resemble a mini QHE trace centred on  $\phi/\phi_o = 1/2$ , consistent with the prediction of recursion in the butterfly spectrum<sup>1</sup>. Inset: the positions of the mini QHE plateaux plotted versus effective magnetic field follow a linear trend (see text).

moiré superlattice makes it possible to study experimentally the complexity of the Hofstadter energy spectrum and, in particular, to investigate the role of electron interactions within it.

## METHODS SUMMARY

Using a co-lamination mechanical transfer technique similar to that described previously<sup>25,26</sup>, graphene–hBN stacks were fabricated on doped Si substrates with a ~300-nm oxide layer. Each stack was then etched into a Hall bar with evaporated Cr/Pd/Au leads using standard electron-beam lithography processes. The bottom boron nitride layer was chosen to be between 10 and 20 nm thick. More than 20 devices were made in this way, with six devices showing similar behaviour to that reported here. We focus only on two high-quality devices in the text, listing other examples in Supplementary Information. The device shown in Fig. 1b represents a typical Hall bar geometry with a channel width of 1 µm and spacing between voltage probes varying from 1 to ~4 µm. Identical transport features were measured across all voltage probes, indicating that the moiré pattern exhibits good uniformity over the entire length of the device. Mobility in our devices ranged from approximately 10,000 to 100,000 cm<sup>2</sup> V<sup>-1</sup> s<sup>-1</sup>, as estimated by fitting a Boltzmann model to the linear response. Four-terminal transport measurements were performed using a lock-in amplifier at 17 Hz with a 10–100-nA source current, using the doped silicon substrate to gate the channel. Samples were measured in a 31-T resistive magnet and a <sup>3</sup>He cryostat (sample in vapour). Longitudinal and Hall conductivities were calculated from the measured resistances according to  $\sigma_{xx} = \rho_{xx} / (\rho_{xx}^2 + R_{xy}^2)$  and  $\sigma_{xy} = R_{xy} / (\rho_{xx}^2 + R_{xy}^2)$ , respectively.

AFM images of the device were acquired at room temperature, using an Omicron low-temperature AFM system. Imaging was performed using a bias voltage of  $V_{\text{bias}} = 0.2$  V and frequency shift of  $\delta f = 20$  Hz. Images were filtered to remove noise.

Received 13 December 2012; accepted 3 April 2013.

Published online 15 May 2013.

- Hofstadter, D. Energy levels and wave functions of Bloch electrons in rational and irrational magnetic fields. *Phys. Rev. B* **14**, 2239–2249 (1976).
- Wannier, G. H. A result not dependent on rationality for Bloch electrons in a magnetic field. *Phys. Status Solidi B* **88**, 757–765 (1978).
- Štředa, P. Quantised Hall effect in a two-dimensional periodic potential. *J. Phys. C* **15**, L1299–L1303 (1982).
- Thouless, D., Kohmoto, M., Nightingale, M. & den Nijs, M. Quantized Hall conductance in a two-dimensional periodic potential. *Phys. Rev. Lett.* **49**, 405–408 (1982).
- MacDonald, A. Landau-level subband structure of electrons on a square lattice. *Phys. Rev. B* **28**, 6713–6717 (1983).
- Koshino, M. & Ando, T. Hall plateau diagram for the Hofstadter butterfly energy spectrum. *Phys. Rev. B* **73**, 155304 (2006).
- Nemec, N. & Cuniberti, G. Hofstadter butterflies of bilayer graphene. *Phys. Rev. B* **75**, 201404(R) (2007).
- Bistritzer, R. & MacDonald, A. Moiré butterflies in twisted bilayer graphene. *Phys. Rev. B* **84**, 035440 (2011).
- Moon, P. & Koshino, M. Energy spectrum and quantum Hall effect in twisted bilayer graphene. *Phys. Rev. B* **85**, 195458 (2012).
- Beugeling, W., Goldman, N. & Morais Smith, C. Topological phases in a two-dimensional lattice: magnetic field versus spin-orbit coupling. *Phys. Rev. B* **86**, 075118 (2012).
- Gerhardts, R., Weiss, D. & Wulf, U. Magnetoresistance oscillations in a grid potential: indication of a Hofstadter-type energy spectrum. *Phys. Rev. B* **43**, 5192–5195 (1991).
- Nakamura, Y., Inoshita, T. & Sakaki, H. Novel magneto-resistance oscillations in laterally modulated two-dimensional electrons with 20nm periodicity formed on vicinal GaAs (111)B substrates. *Physica E* **2**, 944–948 (1998).
- Albrecht, C. *et al.* Fermiology of two-dimensional lateral superlattices. *Phys. Rev. Lett.* **83**, 2234–2237 (1999).
- Schlösser, T., Ensslin, K., Kotthaus, J. P. & Holland, M. Landau subbands generated by a lateral electrostatic superlattice: chasing the Hofstadter butterfly. *Semicond. Sci. Technol.* **11**, 1582–1585 (1996).
- Albrecht, C. *et al.* Evidence of Hofstadter's fractal energy spectrum in the quantized Hall conductance. *Phys. Rev. Lett.* **86**, 147–150 (2001).
- Geisler, M. C. *et al.* Detection of Landau band coupling induced rearrangement of the Hofstadter butterfly. *Physica E* **25**, 227–232 (2004).
- Melinte, S. *et al.* Laterally modulated 2D electron system in the extreme quantum limit. *Phys. Rev. Lett.* **92**, 036802 (2004).
- Harper, P. G. Single band motion of conduction electrons in a uniform magnetic field. *Proc. Phys. Soc. A* **68**, 874–878 (1955).
- von Klitzing, K., Dorda, G. & Pepper, M. New method for high-accuracy determination of the fine-structure constant based on quantized Hall resistance. *Phys. Rev. Lett.* **45**, 494–497 (1980).
- Kuhl, U. & Stöckmann, H. J. Microwave realization of the Hofstadter butterfly. *Phys. Rev. Lett.* **80**, 3232–3235 (1998).
- Jaksch, D. & Zoller, P. Creation of effective magnetic fields in optical lattices: the Hofstadter butterfly for cold neutral atoms. *N. J. Phys.* **5**, 56 (2003).
- Xue, J. *et al.* Scanning tunnelling microscopy and spectroscopy of ultra-flat graphene on hexagonal boron nitride. *Nature Mater.* **10**, 282–285 (2011).
- Decker, R. *et al.* Local electronic properties of graphene on a BN substrate via scanning tunneling microscopy. *Nano Lett.* **11**, 2291–2295 (2011).
- Yankowitz, M. *et al.* Emergence of superlattice Dirac points in graphene on hexagonal boron nitride. *Nature Phys.* **8**, 382–386 (2012).
- Dean, C. R. *et al.* Boron nitride substrates for high-quality graphene electronics. *Nature Nanotechnol.* **5**, 722–726 (2010).
- Dean, C. R. *et al.* Multicomponent fractional quantum Hall effect in graphene. *Nature Phys.* **7**, 693–696 (2011).
- Li, G. *et al.* Observation of Van Hove singularities in twisted graphene layers. *Nature Phys.* **6**, 109–113 (2010).
- Wallbank, J. R., Patel, A. A., Mucha-Kruczynski, M., Geim, A. K. & Fal'ko, V. I. Generic miniband structure of graphene on a hexagonal substrate. Preprint at <http://arxiv.org/abs/1211.4711> (2012).
- Giovannetti, G., Khomyakov, P., Brocks, G., Kelly, P. & Van Den Brink, J. Substrate-induced band gap in graphene on hexagonal boron nitride: ab initio density functional calculations. *Phys. Rev. B* **76**, 073103 (2007).
- Ramasubramaniam, A., Naveh, D. & Towle, E. Tunable band gaps in bilayer graphene-BN heterostructures. *Nano Lett.* **11**, 1070–1075 (2011).

Supplementary Information is available in the online version of the paper.

**Acknowledgements** We thank A. MacDonald for discussions. A portion of this work was performed at the National High Magnetic Field Laboratory, which is supported by US National Science Foundation cooperative agreement no. DMR-0654118, the State of Florida and the US Department of Energy. This work is supported by AFOSR MURI. J.K. and M.I. were supported by the US National Science Foundation under grant no. 0955625. K.L.S. was supported by DARPA under Office of Naval Research contract N00014-1210814. P.K. and F.G. acknowledge sole support from the US Department of Energy (DE-FG02-05ER46215).

**Author Contributions** C.R.D., P. Maher, L.W., C.F., F.G. and Y.G. performed device fabrication and transport measurements. J.K. and M.I. performed AFM measurements. P. Moon and M.K. provided theoretical support. K.W. and T.T. synthesized the hBN samples. K.L.S., J.H. and P.K. advised on experiments. C.R.D., P. Maher, P. Moon, M.K., J.H. and P.K. wrote the manuscript in consultation with all other authors.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.K. (pk2015@columbia.edu).



# Persistent export of $^{231}\text{Pa}$ from the deep central Arctic Ocean over the past 35,000 years

Sharon S. Hoffmann<sup>1</sup>, Jerry F. McManus<sup>1</sup>, William B. Curry<sup>2,3</sup> & L. Susan Brown-Leger<sup>2</sup>

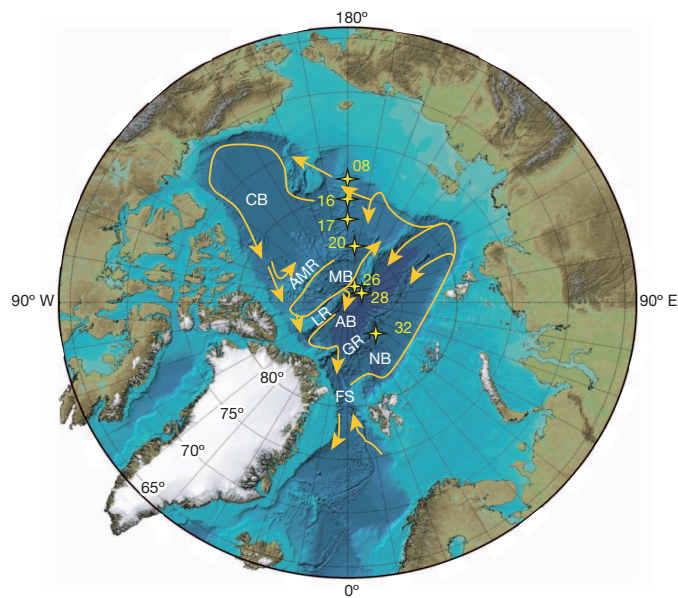
The Arctic Ocean has an important role in Earth's climate, both through surface processes<sup>1</sup> such as sea-ice formation and transport, and through the production and export of waters at depth that contribute to the global thermohaline circulation<sup>2,3</sup>. Deciphering the deep Arctic Ocean's palaeo-oceanographic history is a crucial part of understanding its role in climatic change. Here we show that sedimentary ratios of the radionuclides thorium-230 ( $^{230}\text{Th}$ ) and protactinium-231 ( $^{231}\text{Pa}$ ), which are produced in sea water and removed by particle scavenging on timescales of decades to centuries, respectively<sup>4</sup>, record consistent evidence for the export of  $^{231}\text{Pa}$  from the deep Arctic and may indicate continuous deep-water exchange between the Arctic and Atlantic oceans throughout the past 35,000 years. Seven well-dated box-core records provide a comprehensive overview of  $^{231}\text{Pa}$  and  $^{230}\text{Th}$  burial in Arctic sediments during glacial, deglacial and interglacial conditions. Sedimentary  $^{231}\text{Pa}/^{230}\text{Th}$  ratios decrease nearly linearly with increasing water depth above the core sites, indicating efficient particle scavenging in the upper water column and greater influence of removal by lateral transport at depth. Although the measured  $^{230}\text{Th}$  burial is in balance with its production in Arctic sea water, integrated depth profiles for all time intervals reveal a deficit in  $^{231}\text{Pa}$  burial that can be balanced only by lateral export in the water column. Because no enhanced sink for  $^{231}\text{Pa}$  has yet been found in the Arctic, our records suggest that deep-water exchange through the Fram strait may export  $^{231}\text{Pa}$ . Such export may have continued for the past 35,000 years, suggesting a century-scale replacement time for deep waters in the Arctic Ocean since the most recent glaciation and a persistent contribution of Arctic waters to the global ocean circulation.

Waters in the Arctic are freshened by precipitation and run-off, leading to a net shallow freshwater transport into the Nordic seas<sup>1</sup>. Sea-ice production and melting freshen surface waters through salt rejection, and in some marginal settings also produce cold, saline waters that descend into the Arctic basins<sup>2,3</sup>. These dense waters travel in topographically steered deep currents and eventually transit the Fram strait, the single deep Arctic passage, into the Nordic seas, where they contribute to the production of North Atlantic Deep Water, a major component of the global thermohaline circulation<sup>1–3</sup>. Beneath the cold, fresh surface, warm, salty, Atlantic-sourced waters fill the Arctic depths above 1 km; these waters may have deepened to dominate Arctic intermediate depths during past colder periods as a result of changes in stratification<sup>5</sup>. Such past changes in water masses may imply changing rates of ventilation and circulation.

Although variations in ratios of  $^{231}\text{Pa}$  and  $^{230}\text{Th}$  in sediment can provide information about past rates of oceanographic processes including deep circulation<sup>6–8</sup>, these tracers have seldom been applied in Arctic sediments. Measurements of  $^{231}\text{Pa}$  and  $^{230}\text{Th}$  in the modern-day Arctic water column<sup>9–13</sup> suggest variable particle scavenging rates of  $^{231}\text{Pa}$  and  $^{230}\text{Th}$  in different parts of the basin, including relatively high scavenging rates within the Eurasian basin and in parts of the Canadian basin. The burial rate of  $^{230}\text{Th}$  in central Arctic sediments roughly balances the water column production rate, supporting the use

of  $^{231}\text{Pa}_{\text{xs}}/^{230}\text{Th}_{\text{xs}}$  (where the subscripts refer to 'excess' unsupported isotope activity age-corrected for radioactive decay since deposition; see Supplementary Information) as a tracer of past oceanographic processes in the Arctic<sup>14</sup>. Measured  $^{231}\text{Pa}_{\text{xs}}/^{230}\text{Th}_{\text{xs}}$  ratios in core-top sediments within all four Arctic sub-basins and along the Alaska–Chukchi margin<sup>11,15</sup> fall below the production ratio of 0.093, suggesting a net export of  $^{231}\text{Pa}$ . Two records of  $^{231}\text{Pa}_{\text{xs}}$  from the Arctic provide the first down-core evidence<sup>16,17</sup> for past  $^{231}\text{Pa}$  burial, although neither is sufficiently resolved to give detailed information on patterns of  $^{231}\text{Pa}$  transport and deposition.

Seven box cores (Fig. 1 and Supplementary Table 1) with well-defined radiocarbon chronologies<sup>18,19</sup> (Supplementary Table 2) from the 1994 Arctic Ocean Section cruise provide a cross-Arctic transect at depths within intermediate and deep waters (depths of 1–3.5 km), which we use to examine regional and depth-related trends in  $^{231}\text{Pa}$  deposition and export. Deep-water residence times in these basins can vary from ~150 to 400 yr (refs 20, 21), similar to or longer than the residence times of dissolved  $^{231}\text{Pa}$  in these deep waters before scavenging by particles<sup>11,12</sup>.



**Figure 1 | Locations of seven box cores in a transect across the Arctic Ocean.** PL-94-AR box cores 08, 16, 17 and 20, Makarov basin, 1–3.1-km water depth. PL-94-AR box core 26, Lomonosov ridge, 1 km; box core 28, Amundsen basin, 2 km; box core 32, Nansen basin, 3.5 km. Arctic bathymetric features: CB, Canada basin; AMR, Alpha-Mendeleev ridge; MB, Makarov basin; LR, Lomonosov ridge; AB, Amundsen basin; GR, Gakkel ridge; NB, Nansen basin; FS, Fram strait. For core coordinates and water depths, see Supplementary Table 1. Orange arrows show patterns of intermediate and deep circulation<sup>1–3</sup>. The base map used is the International Bathymetric Chart of the Arctic Ocean.

<sup>1</sup>Lamont-Doherty Earth Observatory of Columbia University, Palisades, New York 10964, USA. <sup>2</sup>Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA. <sup>3</sup>Bermuda Institute of Ocean Sciences, St George's GE 01, Bermuda.

Ratios of  $^{231}\text{Pa}_{\text{xs}}/^{230}\text{Th}_{\text{xs}}$  (Supplementary Table 3) in our late-glacial core sections (here defined as dating from 35 to 15 kyr ago, a period that ended with the start of North Atlantic Bølling–Allerød warming) are almost uniformly below the  $^{231}\text{Pa}/^{230}\text{Th}$  seawater production ratio of 0.093 (Fig. 2), indicating export of  $^{231}\text{Pa}$  from all studied sites. The lowest ratios are seen in the deepest parts of box cores 16 and 17, which are more than 35 kyr old; these samples have large uncertainties in their modelled ages, and the low ratios in these oldest sediments may reflect under-correction for radioactive decay since deposition. There is an increase in  $^{231}\text{Pa}_{\text{xs}}/^{230}\text{Th}_{\text{xs}}$  at the deglaciation (here defined as 15–10 kyr ago) in five cores. In the deepest, box core 32, deglacial ratios remain low; in box core 26, at 1-km depth atop the Lomonosov ridge, ratios increase to the value of the seawater production ratio before 16.5 kyr ago. The  $^{231}\text{Pa}_{\text{xs}}/^{230}\text{Th}_{\text{xs}}$  ratios reach the production ratio, indicating no local export of  $^{231}\text{Pa}$ , in the deglacial intervals in the two other shallowest cores (box cores 08 and 16, at depths of 1–1.6 km) and remain at the production ratio through the Holocene epoch in box cores 08, 16 and 26.

Holocene ratios in intermediate-depth cores vary between near-production-ratio levels similar to those in shallower cores and low ratios similar to those in deeper cores. Box core 17, at a depth of 2.2 km in the Makarov basin, and box core 28, at a depth of 2 km in the Amundsen basin, show synchronous millennial-scale variability in the deglacial and the early Holocene, with the highest peaks in  $^{231}\text{Pa}_{\text{xs}}/^{230}\text{Th}_{\text{xs}}$  occurring ~11 kyr ago, followed by peaks ~7.5 kyr ago (Fig. 2). Surface-driven records such as vertical particle flux (Supplementary Fig. 1 and Supplementary Table 3) do not follow these  $^{231}\text{Pa}_{\text{xs}}/^{230}\text{Th}_{\text{xs}}$  variations, and differ between basins, suggesting that  $^{231}\text{Pa}_{\text{xs}}/^{230}\text{Th}_{\text{xs}}$  is more likely to be controlled by processes at depth such as ventilation and deep advection. The Arctic boundary current splits at the Lomonosov ridge, sending flows into the Amundsen basin and along the slope into the Makarov basin<sup>1</sup>; our sites at similar depths beneath this current in each basin may reflect its influence.

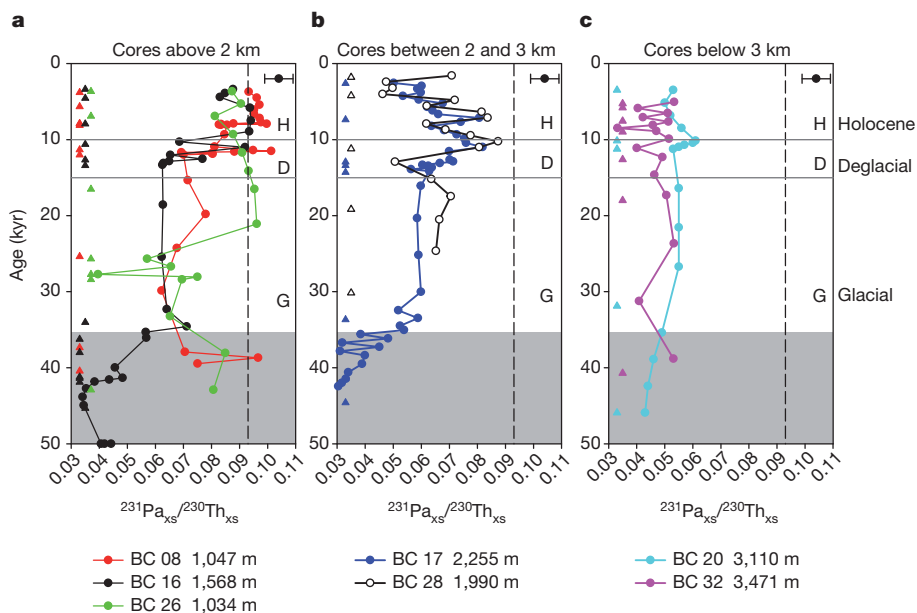
Average  $^{231}\text{Pa}_{\text{xs}}/^{230}\text{Th}_{\text{xs}}$  ratios for the late glacial, deglacial and Holocene show a strong linear relationship with water depth (Fig. 3). During each interval, ratios were highest in shallower (1–1.6 km) cores and lowest in deeper (>3 km) cores. The slope of this relationship is greater during the Holocene and deglacial than during the glacial,

owing to greater differences in shallow ratios between these times. Although increased particle fluxes could lead to increased scavenging of  $^{231}\text{Pa}$  and, thus, higher ratios, our observed average ratio–depth relationship is not accompanied by any apparent depth trend in  $^{230}\text{Th}$ -normalized sediment mass fluxes (Supplementary Fig. 2); nor do particle fluxes seem to be a dominant control on  $^{231}\text{Pa}_{\text{xs}}/^{230}\text{Th}_{\text{xs}}$  in individual cores (Supplementary Fig. 3).

This depth relationship, which has characterized central Arctic  $^{231}\text{Pa}_{\text{xs}}/^{230}\text{Th}_{\text{xs}}$  over the past 35 kyr, mirrors a depth relationship seen in surface sediments of other Arctic cores above 3.5-km water depth<sup>11,15–17</sup> (Supplementary Fig. 4). Core-top  $^{231}\text{Pa}_{\text{xs}}/^{230}\text{Th}_{\text{xs}}$  measurements from 3.5–4.5-km water depth in the Eurasian basin<sup>11</sup> show ratios higher than predicted by our depth trend, although all but one still fall below the production ratio.

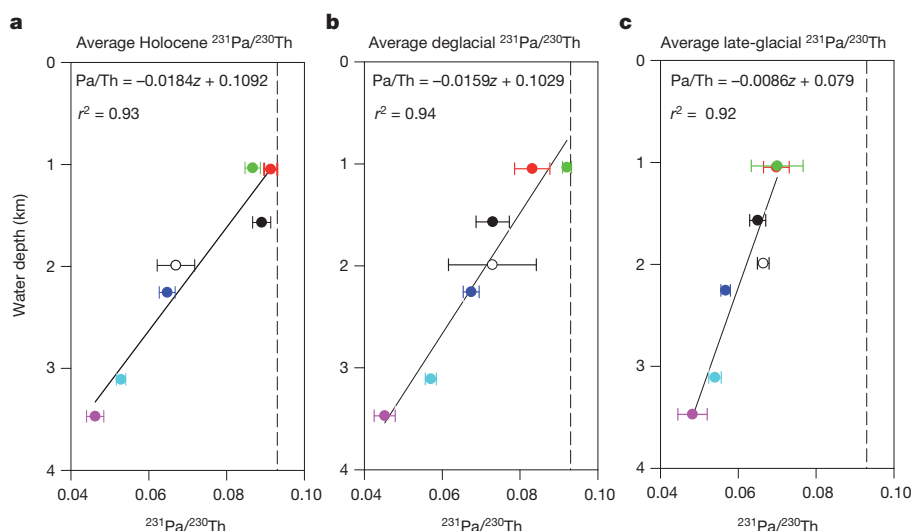
The relationship between  $^{231}\text{Pa}_{\text{xs}}/^{230}\text{Th}_{\text{xs}}$  and water depth that we find in the Arctic Ocean echoes patterns found outside the Arctic despite major dissimilarities in oceanographic setting. It resembles that constrained by  $^{231}\text{Pa}_{\text{xs}}/^{230}\text{Th}_{\text{xs}}$  surface sediments in other oceans including the eastern South Atlantic Ocean<sup>22</sup>, the North Atlantic Ocean<sup>8</sup> and the Pacific Ocean<sup>23</sup>, suggesting that this relationship may be global and related to seawater ingrowth and particle scavenging, rather than to local oceanography. The specific factors responsible for a relationship between sediment  $^{231}\text{Pa}_{\text{xs}}/^{230}\text{Th}_{\text{xs}}$  and water depth remain unclear. Opal scavenges Pa more efficiently than do lithogenic or carbonate particles<sup>24,25</sup> and has been implicated in the ratio–depth relationship elsewhere<sup>22</sup>. Opal contents of samples from our Arctic cores (Supplementary Table 4) show little variation and no clear relation either to water depth or to  $^{231}\text{Pa}_{\text{xs}}/^{230}\text{Th}_{\text{xs}}$  in the same samples, although this may be due to post-depositional dissolution on the sea floor. A simple reversible scavenging–mixing model<sup>26</sup> suggests that, owing to lower activities of  $^{231}\text{Pa}$  at shallower depths, less time is required for shallower waters to reach a steady state with respect to ingrowth and scavenging of  $^{231}\text{Pa}$ . Shorter residence times of  $^{231}\text{Pa}$  in water bathing shallower core sites may lead to more balanced scavenging between  $^{231}\text{Pa}$  and  $^{230}\text{Th}$ , which typically has a short residence time in sea water.

No sink for the  $^{231}\text{Pa}$  exported from our study sites has yet been found in the Arctic. Boundary scavenging, that is, the preferential



**Figure 2** |  $^{231}\text{Pa}_{\text{xs}}/^{230}\text{Th}_{\text{xs}}$  activity ratio records from Arctic box cores. Cores are grouped by water depth: <2 km (a), 2–3 km (b), >3 km (c). Age models are based on radiocarbon dates; ratios are corrected for decay since deposition. Black points with error bars in the upper right corner of each box depict average range of  $1\sigma$  uncertainty, calculated from the standard deviation from the mean

of replicate  $^{231}\text{Pa}_{\text{xs}}/^{230}\text{Th}_{\text{xs}}$  measurements propagated with age uncertainty. Calibrated radiocarbon dates are shown at left in each plot (triangles). The dashed vertical lines show the production ratio of  $^{231}\text{Pa}/^{230}\text{Th}$  in sea water. Grey shading indicates points older than 35 kyr. BC, box core.



**Figure 3 | Average Holocene, deglacial and late glacial  $^{231}\text{Pa}_x/^{230}\text{Th}_x$  activity ratios.** Ratios shown in Fig. 2 were averaged for each core over the Holocene (0–10 kyr ago; **a**), deglacial (10–15 kyr ago; **b**) and late glacial (15–35 kyr ago; **c**). The dashed lines indicate a seawater production ratio of 0.093. Each core is coloured as in Fig. 2. Ratios decrease nearly linearly with increasing

depth during all three climatic periods investigated, although the gradient of this decrease is smaller during the glacial (owing to lower ratios at shallow sites) than during the deglacial and Holocene. Error bars, 1 s.e.;  $r^2$ , coefficient of determination;  $z$ , water depth.

removal at higher-particle-flux margins of  $^{231}\text{Pa}$  imported from lower-particle-flux central basins, has been considered a likely mechanism of  $^{231}\text{Pa}$  removal from the Arctic water column<sup>11,15,27</sup>. Models of particle-reactive metal transport in the ocean support this proposition and suggest that  $^{230}\text{Th}$  may also be subject to boundary scavenging in the most particle-poor parts of the Arctic<sup>27</sup>. Because the existing sedimentary evidence points to local removal of  $^{230}\text{Th}$  from sea water to sediments in the Makarov basin<sup>14</sup>, rather than lateral export of  $^{230}\text{Th}$  to margins, boundary scavenging of  $^{231}\text{Pa}$  should lead to  $^{231}\text{Pa}_x/^{230}\text{Th}_x$  ratios greater than the production ratio at Arctic continental slopes. At the Alaska–Chukchi continental slope in the Canada basin, in which water residence times are greater than 600 yr (ref. 21), and near terrestrial particle sources, core-top ratios are generally less than the production ratio<sup>15</sup> (Supplementary Information), indicating export of  $^{231}\text{Pa}$  produced at depth<sup>15</sup> that is unbalanced within the Arctic proper.

The export of  $^{231}\text{Pa}$  out of the Arctic<sup>15</sup> could also explain low Central Arctic  $^{231}\text{Pa}_x/^{230}\text{Th}_x$  ratios. The Fram strait is the only conduit through which deep water enters and exits the Arctic, and has a sill depth of ~2,500 m. No more than 10% of  $^{230}\text{Th}$  and roughly 39% of  $^{231}\text{Pa}$  produced in the Arctic Ocean is at present exported in water masses exiting southwards through the Fram strait<sup>14,15</sup>. Our results are consistent with a persistent net export of  $^{231}\text{Pa}$ , ranging from 30% in the Holocene to 40% during the Last Glacial Maximum (Methods). Persistent  $^{231}\text{Pa}$  export is also consistent with continuous glacial inflow of warm Atlantic waters reaching intermediate depths during cold intervals<sup>5</sup>, which may have provided salt and mass balance for this deep outflow. Because sedimentary  $^{231}\text{Pa}_x/^{230}\text{Th}_x$  integrates overlying influences<sup>7</sup>, the inferred deepening<sup>5</sup> of an inflow of low- $^{231}\text{Pa}$ , low- $^{230}\text{Th}$  water, such as that from the Atlantic, might also have an impact on the burial of  $^{231}\text{Pa}_x/^{230}\text{Th}_x$ , especially at shallower and intermediate depths.

Increased  $^{231}\text{Pa}$  export during the glacial might also be explained by even greater deep-water exchange through the Fram strait than exists today, although it more probably reflects longer residence times for dissolved Arctic  $^{231}\text{Pa}$  before removal by scavenging, and possibly a combination of these factors. Indeed, it seems likely that heavy perennial sea ice in the glacial Arctic led to diminished particle scavenging and, hence, an increase in dissolved  $^{231}\text{Pa}$  concentrations in the water column, resulting in higher  $^{231}\text{Pa}$  concentrations in water exiting the

Arctic. Low sedimentation rates and  $^{230}\text{Th}$ -normalized sediment mass fluxes (Supplementary Information), as well as reductions in metal-liferous deposition<sup>28</sup>, in glacial-age Arctic sediments are consistent with this hypothesis. The ~40% decrease in glacial scavenging intensity suggested by lower particle fluxes (Supplementary Information) could counter the ~33% increase in  $^{231}\text{Pa}$  export, to result in a small net reduction (<10%) in glacial deep-water outflow through the Fram strait, although the uncertainties associated with these estimates are too large to resolve a change of this scale.

The eventual fate of the  $^{231}\text{Pa}$  exported from the Arctic is presently unknown. Although  $^{231}\text{Pa}$  burial could occur farther south in the Nordic seas or North Atlantic, the Fram strait, with higher particle fluxes due to seasonal productivity, sea-ice transport and nearby terrigenous particle sources, could act as a likely sink for the  $^{231}\text{Pa}$  exported from the Arctic.

The combined observations of a balanced Arctic  $^{230}\text{Th}$  budget and persistent sedimentary  $^{231}\text{Pa}_x$  deficits provide constraints on the time-scale of  $^{231}\text{Pa}$  removal from the basin. Residence times with respect to removal by scavenging for these nuclides in the water column in other oceans (10–20 yr for  $^{230}\text{Th}$ , 50–200 yr for  $^{231}\text{Pa}$ ; refs 29, 30) indicate that the deep waters of the Arctic must be exchanged through the Fram strait on centennial timescales. Especially given the evidence for variable and possibly longer residence times in sea water for  $^{231}\text{Pa}$  and  $^{230}\text{Th}$  in the Arctic<sup>9–13</sup>, any decadal-scale exchange would lead to sedimentary  $^{230}\text{Th}_x$  deficits that are not observed<sup>14</sup>. Longer, millennial, timescales of exchange would not allow the widespread net sedimentary  $^{231}\text{Pa}_x$  deficit evident in core tops<sup>11,15–17</sup> and our new down-core results. Although our evidence for greater glacial export of  $^{231}\text{Pa}$  is more likely to reflect diminished scavenging than enhanced deep outflow, the persistent outflow from a presumably perennially ice-covered Arctic suggests that the same sea ice and brine formation processes as today probably occurred in recurring open waters at the basin margins. The century-scale deep-water exchange is consistent with a residence time for seawater in the modern Arctic of 120–360 yr (Methods), assuming a volume of  $11.455 \times 10^6 \text{ km}^3$  and a volume transport through the Fram strait of approximately 1–3 Sv (refs 3, 15). It is also consistent with estimates of the ‘age’ of deep waters in the Arctic basins<sup>20,21</sup>. Because the most important component of the Fram strait exchange for Arctic  $^{231}\text{Pa}$  export is the intermediate and deep outflow, the results presented here provide a proxy for reconstructing past



dynamics of the Arctic Ocean at depth, and indicate that the outflow through the Fram strait may have persisted through the very different conditions and dramatic climate changes since the most recent glacial.

## METHODS SUMMARY

Cores were sampled in 0.5-cm slices every 1 or 0.5 cm. Sediments were spiked with  $^{233}\text{Pa}$  and  $^{229}\text{Th}$  for isotope dilution analysis and digested in  $\text{HClO}_4$ , HF and  $\text{HNO}_3$ . Aliquots were spiked with  $^{236}\text{U}$  and  $^{229}\text{Th}$  for measurement of  $^{238}\text{U}$  and  $^{232}\text{Th}$ . Thorium and Pa fractions were purified by anion-exchange column chromatography and analysed on a Thermo-Finnigan Element 2 inductively coupled plasma mass spectrometer<sup>7,8</sup> (ICP-MS). Measurements of  $^{238}\text{U}$  and  $^{232}\text{Th}$  were used to estimate supported detrital and ingrown  $^{231}\text{Pa}$  and  $^{230}\text{Th}$  activity. Corrections assuming respective detrital  $^{238}\text{U}/^{232}\text{Th}$  ratios of 0.5 and 0.7 were applied to each sample to calculate a range of corrected  $^{231}\text{Pa}_{\text{xs}}$  and  $^{230}\text{Th}_{\text{xs}}$  values and their ratios<sup>7,8</sup>; these ratio values are reported in Supplementary Table 3, and average values based on these ranges are plotted in Fig. 2 and Supplementary Fig. 1. Excess values were also corrected for radioactive decay since deposition and ingrowth from authigenic U estimated using the same values for  $^{238}\text{U}/^{232}\text{Th}$ .

The uncertainty associated with ICP-MS analysis is estimated to be less than 2% for both the  $^{231}\text{Pa}$  and the  $^{230}\text{Th}$  analysis. The uncertainty in our age models produces an uncertainty in age-corrected  $^{231}\text{Pa}_{\text{xs}}/^{230}\text{Th}_{\text{xs}}$  ratios of less than 3% for age errors up to  $\pm 2$  kyr, incorporating uncertainties in  $^{14}\text{C}$  analyses, reservoir correction and calendar age calibration. The average  $1\sigma$  uncertainty in  $^{231}\text{Pa}_{\text{xs}}/^{230}\text{Th}_{\text{xs}}$ , calculated from replicate analyses (Supplementary Table 5) and incorporating age uncertainty, was  $\sim 8\%$ .

The data reported here, and tabulated in Supplementary Information, will be archived in the US NOAA Paleoclimatology database.

**Full Methods** and any associated references are available in the online version of the paper.

Received 11 October 2012; accepted 4 April 2013.

- Rudels, B. in *Encyclopedia of Ocean Sciences* (eds Steele, J. H., Turekian, K. K. & Thorpe, S. A.) 211–225 (Academic, 2009).
- Aagaard, K., Swift, J. H. & Carmack, E. C. Thermohaline circulation in the Arctic Mediterranean Seas. *J. Geophys. Res.* **90**, 4833–4846 (1985).
- Rudels, B., Friedrich, H. J. & Quadfasel, D. The arctic circumpolar boundary current. *Deep Sea Res.* **46**, 1023–1062 (1999).
- Henderson, G. M. & Anderson, R. F. The U-series toolbox for paleoceanography. *Rev. Mineral. Geochem.* **52**, 493–531 (2003).
- Cronin, T. M. *et al.* Deep Arctic Ocean warming during the last glacial cycle. *Nature Geosci.* **5**, 631–634 (2012).
- Yu, E.-F., Francois, R. & Bacon, M. P. Similar rates of modern and last-glacial ocean circulation inferred from radiochemical data. *Nature* **379**, 689–694 (1996).
- McManus, J. F., Francois, R., Gherardi, J.-M., Keigwin, L. D. & Brown-Leger, S. Collapse and rapid resumption of Atlantic meridional circulation linked to deglacial climate changes. *Nature* **428**, 834–837 (2004).
- Gherardi, J.-M. *et al.* Glacial-interglacial circulation changes inferred from  $^{231}\text{Pa}/^{230}\text{Th}$  sedimentary record in the North Atlantic region. *Paleoceanography* **24**, PA2204 (2009).
- Bacon, M. P., Huh, C.-A. & Moore, R. M. Vertical profiles of some natural radionuclides over the Alpha Ridge, Arctic Ocean. *Earth Planet. Sci. Lett.* **95**, 15–22 (1989).
- Cochran, J. K., Hirschberg, D. J., Livingston, H. D., Buesseler, K. O. & Key, R. M. Natural and anthropogenic radionuclide distributions in the Nansen Basin, Arctic Ocean: scavenging rates and circulation timescales. *Deep Sea Res.* **42**, 1495–1517 (1995).
- Scholten, J. C., Rutgers van der Loeff, M. M. & Michel, A. Distribution of  $^{230}\text{Th}$  and  $^{231}\text{Pa}$  in the water column in relation to the ventilation of the deep Arctic basins. *Deep Sea Res.* **42**, 1519–1531 (1995).
- Edmonds, H. N., Moran, S. B., Cheng, H. & Edwards, R. L.  $^{230}\text{Th}$  and  $^{231}\text{Pa}$  in the Arctic Ocean: implications for particle fluxes and basin-scale Th/Pa fractionation. *Earth Planet. Sci. Lett.* **227**, 155–167 (2004).
- Trimble, S. M., Baskaran, M. & Porcelli, D. Scavenging of thorium isotopes in the Canada Basin of the Arctic Ocean. *Earth Planet. Sci. Lett.* **222**, 915–932 (2004).
- Hoffmann, S. S. & McManus, J. F. Is there a  $^{230}\text{Th}$  deficit in Arctic sediments? *Earth Planet. Sci. Lett.* **258**, 516–527 (2007).
- Moran, S. B. *et al.*  $^{231}\text{Pa}$  and  $^{230}\text{Th}$  in surface sediments of the Arctic Ocean: implications for  $^{231}\text{Pa}/^{230}\text{Th}$  fractionation, boundary scavenging, and advective export. *Earth Planet. Sci. Lett.* **234**, 235–248 (2005).
- Ku, T.-L. & Broecker, W. S. Rates of sedimentation in the Arctic Ocean. *Prog. Oceanogr.* **4**, 95–104 (1965).
- Not, C. & Hillaire-Marcel, C. Time constraints from  $^{230}\text{Th}$  and  $^{231}\text{Pa}$  data in late Quaternary, low sedimentation rate sequences from the Arctic Ocean: an example from the northern Mendeleev Ridge. *Quat. Sci. Rev.* **29**, 3665–3675 (2010).
- Poore, R. Z., Ostermann, D. R. & McGehee, J. Stable isotope data and AMS  $^{14}\text{C}$  dates from Arctic Ocean Section 1994 surface sediment transect and box core samples from the Mendeleev Ridge area. Open-File Report 99–48 (USGS, 1999).
- Darby, D. A., Bischof, J. F. & Jones, G. A. Radiocarbon chronology of depositional regimes in the western Arctic Ocean. *Deep Sea Res.* **44**, 1745–1757 (1997).
- Schlosser, P. *et al.* The first trans-Arctic  $^{14}\text{C}$  section: comparison of the mean ages of the deep waters in the Eurasian and Canadian basins of the Arctic Ocean. *Nucl. Instrum. Methods* **123B**, 431–437 (1997).
- Östlund, H. G., Possnert, G. & Swift, J. H. Ventilation rate of the deep Arctic Ocean from carbon 14 data. *J. Geophys. Res.* **92**, 3769–3777 (1987).
- Scholten, J. C. *et al.* Advection and scavenging: effects on  $^{230}\text{Th}$  and  $^{231}\text{Pa}$  distribution off Southwest Africa. *Earth Planet. Sci. Lett.* **271**, 159–169 (2008).
- Lao, Y. *et al.* Transport and burial rates of Be-10 and Pa-231 in the Pacific Ocean during the Holocene period. *Earth Planet. Sci. Lett.* **113**, 173–189 (1992).
- Chase, Z., Anderson, R. F., Fleischer, M. Q. & Kubik, P. W. The influence of particle composition and particle flux on scavenging of Th, Pa and Be in the ocean. *Earth Planet. Sci. Lett.* **204**, 215–229 (2002).
- Siddall, M.  $^{231}\text{Pa}/^{230}\text{Th}$  fractionation by ocean transport, biogenic particle flux and particle type. *Earth Planet. Sci. Lett.* **237**, 135–155 (2005).
- Luo, Y., Francois, R. & Allen, S. E. Sediment  $^{231}\text{Pa}/^{230}\text{Th}$  as a recorder of the rate of the Atlantic meridional overturning circulation: insights from a 2-D model. *Ocean Sci. Discuss.* **6**, 382–400 (2010).
- Roy-Barman, M. Modeling the effect of boundary scavenging on thorium and protactinium profiles in the ocean. *Biogeosciences* **6**, 3091–3107 (2009).
- März, C., Poulton, S. W., Brumsack, H.-J. & Wagner, T. Manganese-rich brown layers in Arctic Ocean sediments: composition, formation mechanisms, and diagenetic overprint. *Geochim. Cosmochim. Acta* **75**, 7668–7687 (2011).
- Anderson, R. F., Bacon, M. P. & Brewer, P. G. Removal of  $^{230}\text{Th}$  and  $^{231}\text{Pa}$  from the open ocean. *Earth Planet. Sci. Lett.* **62**, 7–23 (1983).
- Yu, E.-F., Francois, R., Bacon, M. P. & Fleer, A. P. Fluxes of  $^{230}\text{Th}$  and  $^{231}\text{Pa}$  to the deep sea: implications for the interpretation of excess  $^{230}\text{Th}$  and  $^{231}\text{Pa}/^{230}\text{Th}$  profiles in sediments. *Earth Planet. Sci. Lett.* **191**, 219–230 (2001).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank R. Poore for sharing radiocarbon dates for box cores 28 and 32. We thank O. Marchal, R. Newton, L. F. Robinson, D. McCorkle, E. Boyle, H. Oates and M. Jackson for discussions. The manuscript was improved by the comments of A. Thomas. This work was supported in part by the Comer Science and Education Fund and by US NSF grants OCE-0402565 (J.F.M.), OCE-0550637 (J.F.M.), OCE-0902985 (J.F.M.), AGS-0936496 (J.F.M.) and ARC-0520073 (W.B.C.).

**Author Contributions** S.S.H. and J.F.M. designed the study with input from W.B.C. S.S.H. carried out core sampling, sediment digestions and instrumental analyses, and wrote the majority of the paper. S.B.-L. contributed to sediment digestion and analytical laboratory procedures. J.F.M. contributed to writing and revising the paper. All authors discussed the results and commented on the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.S.H. (ssh13@columbia.edu).

## METHODS

**Sampling and isotopic analysis.** Cores were sampled in 0.5-cm slices every 1 cm (all box cores except 17) or 0.5 cm (box core 17). Sediments were spiked with  $^{233}\text{Pa}$  and  $^{229}\text{Th}$  for isotope dilution analysis and digested in  $\text{HClO}_4$ ,  $\text{HF}$  and  $\text{HNO}_3$ . Aliquots were spiked with  $^{236}\text{U}$  and  $^{229}\text{Th}$  for measurement of  $^{238}\text{U}$  and  $^{232}\text{Th}$ . Thorium and Pa fractions were purified by anion-exchange column chromatography and analysed on a Thermo-Finnigan Element 2 ICP-MS<sup>31</sup>. Repeated measurements of  $^{235}\text{U}/^{238}\text{U}$  ratios in the U standard solution NBS960 were used to correct for isotope mass fractionation during ICP-MS analysis. Measurements of  $^{238}\text{U}$  and  $^{232}\text{Th}$  were used to estimate supported detrital and ingrown  $^{231}\text{Pa}$  and  $^{230}\text{Th}$  activity. Corrections assuming respective detrital  $^{238}\text{U}/^{232}\text{Th}$  ratios of 0.5 and of 0.7 were applied to each sample to calculate a range of corrected  $^{231}\text{Pa}_{\text{xs}}$  and  $^{230}\text{Th}_{\text{xs}}$  values and their ratios; these ratio values are reported in Supplementary Table 3. Average values based on the corrected ranges are reported in Supplementary Table 3 and plotted in Fig. 2 and Supplementary Fig. 1. Excess values were corrected for effects of both radioactive decay since deposition and authigenic ingrowth. (Published core-top excess ratios from other studies discussed in this paper have not been corrected for decay since deposition.)

The uncertainty associated with ICP-MS analysis is estimated to be less than 2% for both the  $^{231}\text{Pa}$  and the  $^{230}\text{Th}$  analysis. The uncertainty in our age models produces an uncertainty in age-corrected  $^{231}\text{Pa}_{\text{xs}}/^{230}\text{Th}_{\text{xs}}$  ratios of less than 3% for age errors up to  $\pm 2$  kyr. The average  $1\sigma$  uncertainty in  $^{231}\text{Pa}_{\text{xs}}/^{230}\text{Th}_{\text{xs}}$  calculated from replicate analyses was  $\sim 8\%$ .

**Core chronologies.** We constructed age models by linear interpolation between radiocarbon-dated intervals, using both published dates<sup>18,19</sup> and dates new to this study. Radiocarbon ages were converted to calendar years using the CALIB 5.2 program<sup>32</sup> and the MARINECAL 04 marine calibration data set<sup>33</sup> for ages less than 22 kyr, and the calibration from ref. 34 for greater ages<sup>34</sup>. A local difference in reservoir age of  $\Delta R = 40$  yr, to provide a reservoir age of 440 yr, following the studies in which a number of the dates used here were published<sup>18,19,35</sup>, plus an additional 250-yr  $\Delta R$  (refs 36–38), to account for the likely additional influence of sea ice on air–sea gas exchange, were applied to all ages.

Published radiocarbon dates were measured from samples of the polar planktonic foraminifera *Neogloboquadrina pachyderma* sinistral, for box cores 08, 16 and 17<sup>18</sup> (from the same subcores that we sampled for nuclide records). Published dates for box core 20<sup>19</sup> were measured in a different subcore from that used for nuclide sampling. Although stratigraphy can vary by several centimetres between subcores of the same box core, our coarse fraction weight per cent record in box core 20 matches the published record<sup>19</sup> quite well, suggesting little stratigraphic offset between that subcore and ours. Dates for box cores 28 and 32 (R. Poore, personal communication) were produced from the same subcores as our radionuclide records. Dates from box core 26 (this study) were measured from

*N. pachyderma* sinistral at the NOSAMS facility at Woods Hole Oceanographic Institution.

**Additional methods.** Biogenic silica contents (Supplementary Table 4) of selected samples were measured at the Lamont-Doherty Earth Observatory using standard methods<sup>39</sup>.

$^{230}\text{Th}_{\text{xs}}$ -normalized sediment mass fluxes<sup>40</sup> to the seafloor were calculated according to the equation  $F = \beta z / ^{230}\text{Th}_{\text{xs}}$ , where  $\beta$  is the production rate of  $^{230}\text{Th}$  in seawater ( $0.0267 \text{ d.p.m. m}^{-3} \text{ kyr}^{-1}$ ),  $z$  is the water depth of the site (in metres) and  $F$  is reported in units of  $\text{g cm}^{-2} \text{ kyr}^{-1}$ .

Our estimates of persistent Holocene and glacial export of  $^{231}\text{Pa}$  from the Arctic Ocean as a whole were calculated using estimates of Arctic water volume, at 0.5-km depth increments, inferred from Arctic basin hypsometry<sup>41</sup>, and the average  $^{231}\text{Pa}$  export within each such water depth increment was inferred from  $^{231}\text{Pa}_{\text{xs}}/^{230}\text{Th}_{\text{xs}}$  ratios and the production ratio of  $^{231}\text{Pa}$  and  $^{230}\text{Th}$  in sea water. The residence time of sea water within the Arctic (120–360 yr) was calculated by dividing the Arctic basin volume<sup>41</sup> ( $11.455 \times 10^6 \text{ km}^3$ ) by the outflow through the Fram strait<sup>15</sup> ( $1\text{--}3 \text{ Sv}$ ).

- Choi, M. S. *et al.* Rapid determination of  $^{230}\text{Th}$  and  $^{231}\text{Pa}$  in seawater by desolvated micro-nebulization inductively coupled plasma magnetic sector mass spectrometry. *Mar. Chem.* **76**, 99–112 (2001).
- Stuiver, M., Reimer, P. J. & Reimer, R. W. CALIB Radiocarbon Calibration. <http://calib.qub.ac.uk/calib/> (version 5, 2005).
- Hughen, K. A. *et al.* Marine04 marine radiocarbon age calibration, 26-0 ka BP. *Radiocarbon* **46**, 1059–1086 (2004).
- Fairbanks, R. G. *et al.* Marine radiocarbon calibration curve spanning 0 to 50,000 years B.P. based on paired  $^{230}\text{Th}/^{234}\text{U}/^{238}\text{U}$  and  $^{14}\text{C}$  dates on pristine corals. *Quat. Sci. Rev.* **24**, 1781–1796 (2005).
- Poore, R. Z., Osterman, L., Curry, W. B. & Phillips, R. L. Late Pleistocene and Holocene meltwater events in the western Arctic Ocean. *Geology* **27**, 759–762 (1999).
- Mangerud, J. & Gulliksen, S. Apparent radiocarbon ages of recent marine shells from Norway, Spitsbergen, and Arctic Canada. *Quat. Res.* **5**, 263–273 (1975).
- Butzin, M., Prange, M. & Lohmann, G. Radiocarbon simulations for the global ocean: the effects of wind stress, Southern Ocean sea ice and Heinrich events. *Earth Planet. Sci. Lett.* **235**, 45–61 (2005).
- Hanslik, D. *et al.* Quaternary Arctic Ocean sea ice variations and radiocarbon reservoir age corrections. *Quat. Sci. Rev.* **29**, 3430–3441 (2010).
- Mortlock, R. A. & Froelich, P. N. A simple method for the rapid determination of biogenic opal in pelagic marine sediments. *Deep Sea Res. A* **36**, 1415–1426 (1989).
- Francois, R., Frank, M., Rutgers van der Loeff, M. M. & Bacon, M. P.  $^{230}\text{Th}$  normalization: An essential tool for interpreting sedimentary fluxes during the late Quaternary. *Paleoceanography* **19**, PA1018 (2004).
- Jakobsson, M. Hypsometry and volume of the Arctic Ocean and its constituent seas. *Geochem. Geophys. Geosyst.* **3**, <http://dx.doi.org/10.1029/2001GC000302> (2002).

# Emergence of two types of terrestrial planet on solidification of magma ocean

Keiko Hamano<sup>1</sup>, Yutaka Abe<sup>1</sup> & Hidenori Genda<sup>1,2</sup>

Understanding the origins of the diversity in terrestrial planets is a fundamental goal in Earth and planetary sciences. In the Solar System, Venus has a similar size and bulk composition to those of Earth, but it lacks water<sup>1–3</sup>. Because a richer variety of exoplanets is expected to be discovered, prediction of their atmospheres and surface environments requires a general framework for planetary evolution. Here we show that terrestrial planets can be divided into two distinct types on the basis of their evolutionary history during solidification from the initially hot molten state expected from the standard formation model<sup>4,5</sup>. Even if, apart from their orbits, they were identical just after formation, the solidified planets can have different characteristics. A type I planet, which is formed beyond a certain critical distance from the host star, solidifies within several million years. If the planet acquires water during formation, most of this water is retained and forms the earliest oceans. In contrast, on a type II planet, which is formed inside the critical distance, a magma ocean can be sustained for longer, even with a larger initial amount of water. Its duration could be as long as 100 million years if the planet is formed together with a mass of water comparable to the total inventory of the modern Earth. Hydrodynamic escape desiccates type II planets during the slow solidification process. Although Earth is categorized as type I, it is not clear which type Venus is because its orbital distance is close to the critical distance. However, because the dryness of the surface and mantle predicted for type II planets is consistent with the characteristics of Venus, it may be representative of type II planets. Also, future observations may have a chance to detect not only terrestrial exoplanets covered with water ocean but also those covered with magma ocean around a young star.

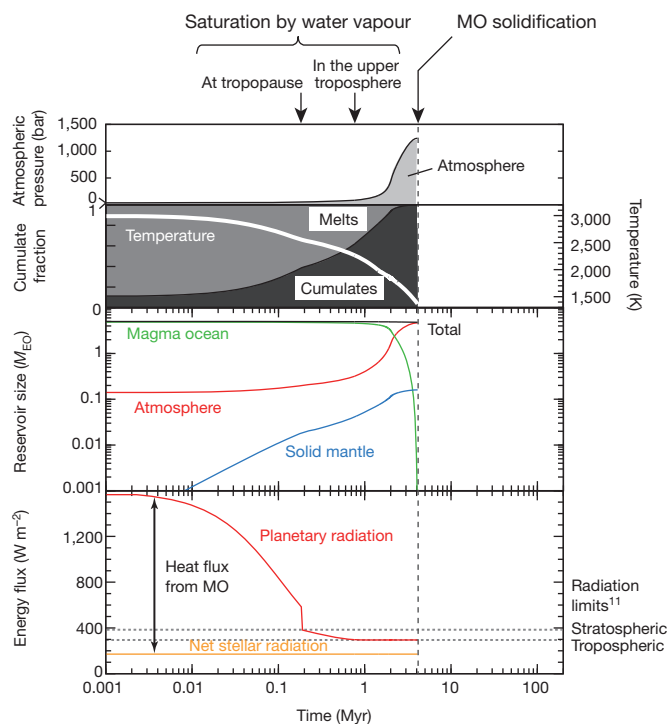
Theoretical studies on planet formation suggest that Earth-sized planets (Earth and Venus) should form as a result of giant impacts between protoplanets<sup>4</sup>, and probably all start their lives in a globally molten state<sup>5</sup>. The earliest phase of planetary evolution is thus solidification of a magma ocean, which provides the initial conditions for mantle differentiation and distribution of volatiles between the interior and the surface<sup>6</sup>. The timing of the end of this phase also determines the starting point for subsequent events such as water ocean formation, and possibly the onset of plate tectonics and the development of life.

The thermal evolution of a magma ocean is closely linked to the formation of a steam atmosphere<sup>6–8</sup>. A massive steam atmosphere decreases outgoing radiation from the planet through its strong greenhouse effects, and thus delays the solidification process. Degassing from a solidifying magma ocean can in turn greatly increase the amount of the steam atmosphere<sup>6</sup>. Therefore, evaluating the outgoing radiation during atmospheric evolution is the key to understanding the earliest thermal history of the planet.

Particularly for a hot steam atmosphere above a magma ocean, previous studies have shown that the outgoing radiation has a lower limit (about  $300 \text{ W m}^{-2}$ ) owing to saturation of the upper troposphere with water vapour<sup>9–11</sup>. This is because its thermal structure is uniquely determined by the saturated-vapour-pressure curve of water (see

ref. 11 for a further explanation). Because the net incoming stellar radiation in the zone of formation of terrestrial planets can be smaller or larger than this radiation limit, it can be speculated from heat balance that the presence of such a limit would produce an evolutionary dichotomy between terrestrial planets. The effect of the radiation limit, however, was neglected in a previously proposed coupled model for a deep magma ocean and steam atmosphere<sup>6</sup>. In addition, the greenhouse effect of water vapour has not been fully considered.

Here we perform radiative–convective equilibrium calculations to enable our model to be used to calculate the radiation limit for a saturated steam atmosphere. In addition we take into account the water loss associated with hydrodynamic escape<sup>12</sup>, which is expected to occur in parallel with magma-ocean solidification. Competition between degassing and water loss determines whether the steam atmosphere grows or escapes. This would affect not only the thermal evolution of planets but also the planetary water inventory at the time of complete solidification.



**Figure 1 | Typical evolution of a type I planet.** Evolution of a planet located at 1 AU, with an initial water mass of five times the current ocean mass on Earth ( $M_{EO}$ ,  $1.4 \times 10^{21} \text{ kg}$ ). The grey dotted lines in the bottom panel indicate the radiation limits<sup>11</sup>. At  $\sim 0.7 \text{ Myr}$ , the planetary radiation reaches the tropospheric radiation limit. After that, the heat flux from the magma ocean (MO) becomes constant. This results in rapid solidification at  $\sim 4 \text{ Myr}$ . The solidification time and final water partitioning are comparable to those reported in ref. 6 for a MO 2,000 km deep on Earth and a total water mass of  $\sim 10 M_{EO}$ .

<sup>1</sup>Department of Earth and Planetary Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. <sup>2</sup>Earth-Life Science Institute, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo 152-8551, Japan.



Figure 1 shows the typical evolution of an Earth-sized planet located at a distance of 1 AU from its parent star. Because of the high solubility of water in silicate melts, a sizeable fraction of the water dissolves in the initial deep magma ocean. As the magma solidifies, degassing of water from the interior leads to an increase in the atmospheric mass. At about 0.2 million years (Myr), because of the growth of the atmosphere combined with the decrease in temperature, the atmosphere starts to be saturated with water vapour at the tropopause. As the saturation front moves deeper into the troposphere, the planetary radiation decreases until it finally becomes equal to the tropospheric radiation limit at about 0.7 Myr. The outgoing radiation flux then remains constant throughout the subsequent solidification period.

The overall solidification time is about 4 Myr in this case, and is mainly determined by the minimum net heat flux; this is defined as the difference between the radiation limit and the net incident stellar radiation. The existence of a minimum heat flux after saturation gives rise to rapid solidification because further atmospheric growth has no effect on the cooling rate of the planet. Hydrodynamic escape makes little contribution to the water inventory during the short magma-ocean period. We refer to this type of planet as type I.

This mechanism does not apply to the evolution of a planet at 0.7 AU (Fig. 2), in which the net incident stellar radiation exceeds the tropospheric radiation limit. In the earlier stages up to  $\sim 0.2$  Myr, its evolution is similar to that of a type I planet. However, at  $\sim 1$  Myr the outgoing radiation almost balances the net incident stellar radiation before a sufficiently deep part of the troposphere becomes saturated. This results in an extremely low heat flux so that the magma ocean is sustained for about 100 Myr, much longer than the value of 4 Myr at 1 AU, for an initial water amount of five times the current ocean mass on Earth ( $5M_{\text{EO}}$ ). During this period there is a large decrease in the total water inventory of the planet.

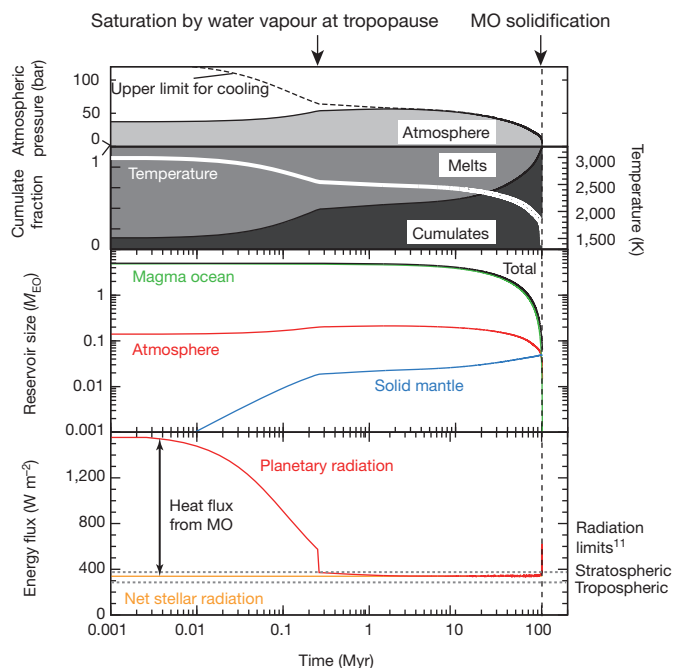
The solidification rate is mainly governed by the water loss rate due to hydrodynamic escape. Because the outgoing radiation must exceed

the incident stellar radiation for cooling, the steam atmosphere must be optically thinner for lower surface temperatures (Fig. 2), so that a net loss of the steam atmosphere is required. We refer to this type of planet as type II.

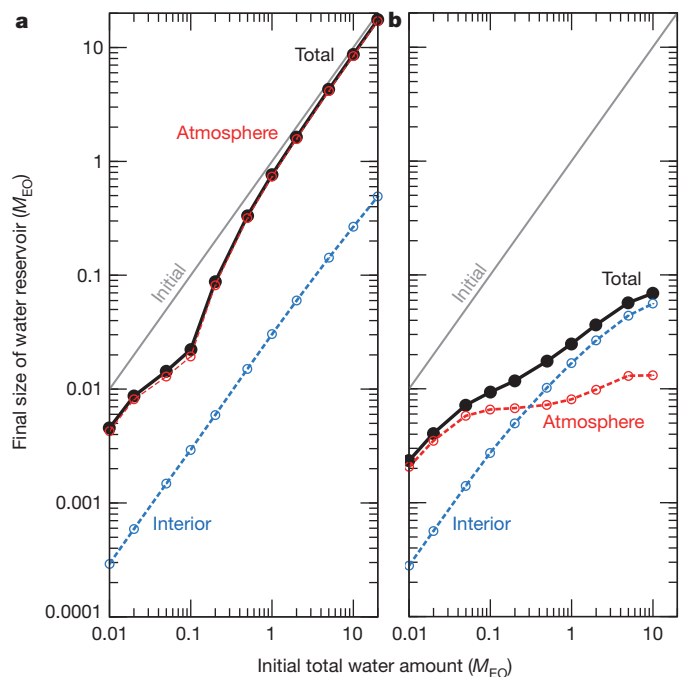
These two types of planet also have a different partitioning of water between the atmosphere and the interior (Fig. 3). For the type I planet at 1 AU, the total water inventory becomes less susceptible to loss for larger initial water masses. The primary water reservoir is the steam atmosphere, whereas the deep interior accounts for only a small percentage of the planetary water inventory. In contrast, the type II planet at 0.7 AU becomes desiccated during the longer magma-ocean period. The final total water inventory is less than  $0.1M_{\text{EO}}$  even for an initial value of  $10M_{\text{EO}}$ . In addition, as the initial water inventory increases, the interior reservoir becomes larger than that of the atmosphere.

Figure 4 shows the solidification time and final total water inventory for Earth-sized planets within the terrestrial planet formation zone. The solidification time (see Supplementary Information for analytical expressions) is seen to peak at about 0.8 AU. Planets located beyond this range are classified as type I, whereas those inside it are classified as type II. Type I planets have maximum solidification times typically as short as several million years owing to the minimum heat flux. Because only a limited amount of hydrodynamic escape can occur within such a short period, it has a reduced impact on the thermal history and water budget of the planet. Soon after solidification, water oceans would probably form on the surface.

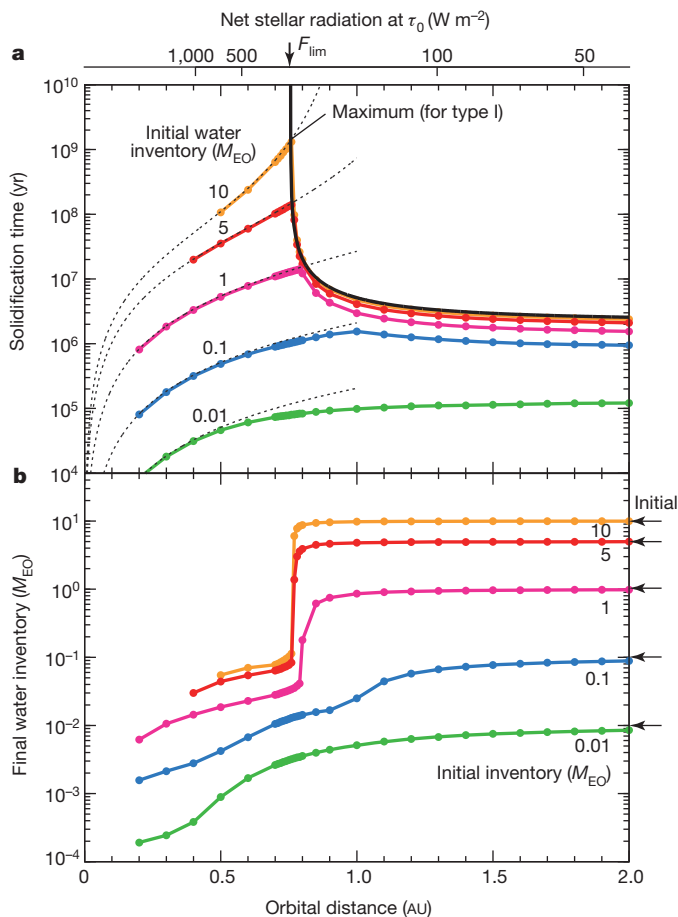
For a type I planet, the initial total water inventory simply affects the volume of the earliest oceans, whereas it has a strong influence on the solidification time for a type II planet. As seen in Fig. 4, the duration of the magma-ocean phase agrees well with the time required for total loss of the primordial water. This reflects the fact that the type II planet must lose water to cool down, and consequently a larger water endowment results in a longer magma-ocean period. The final total water inventory of type II planets never exceeds  $0.1 M_{\text{EO}}$ , irrespective of the initial value.



**Figure 2 | Typical evolution of a type II planet.** As in Fig. 1, but for an orbital distance of 0.7 AU. The net incident stellar radiation is larger than the tropospheric radiation limit. At  $\sim 1$  Myr, planetary radiation almost balances the net stellar radiation. The subsequent heat flux from the MO is controlled by the rate of water loss. The dashed black line in the top panel represents the upper limit of atmospheric pressure for cooling, for which the planetary radiation is equal to the net stellar flux. The MO period is  $\sim 100$  Myr, much longer than that for a type I planet.



**Figure 3 | Water partitioning between steam atmosphere and planetary interior.** The amount of water depends on the rate of loss and the duration of the MO period. **a**, Planet at 1 AU (type I). Most of the primordial water remains and contributes to the steam atmosphere at the time of complete solidification. **b**, Planet at 0.7 AU (type II). The final total mass of water is less than  $0.1M_{\text{EO}}$ . The planetary interior may become the dominant reservoir if water is incorporated before the silicate melts become desiccated. The effect of a surface thermal boundary layer is discussed in Supplementary Information.



**Figure 4 | Two distinct types of terrestrial planet.** The upper x axis shows the corresponding initial net stellar radiation. The arrow indicates the tropospheric radiation limit. The critical orbital distance  $a_{\text{cr}}$  of  $\sim 0.76$  AU separates the orbital regimes of the two types of planet. **a**, Solidification time. The dotted lines show the time required for complete loss of primordial water. This provides a good approximation of the solidification time of type II planets. The maximum solidification time for type I planets is also shown (see Supplementary Information). **b**, Total water inventory at the time of complete solidification. A strong transition is exhibited at about  $a_{\text{cr}}$ .

The different mechanisms determining the solidification rate thus also lead to a sharp transition in the planetary water budget.

Whether a type I or a type II planet is formed around a star depends on the critical orbital distance  $a_{\text{cr}}$ , which is defined as the distance at which the net stellar flux that the planet receives equals the tropospheric radiation limit  $F_{\text{lim}}$  of the steam atmosphere,

$$a_{\text{cr}} \approx 0.76 \left( \frac{F_{\text{lim}}}{294 \text{ W m}^{-2}} \right)^{-1/2} \left( \frac{S_{\text{stl}}(\tau_0)}{0.7 S_{\odot}} \right)^{1/2} \left( \frac{1-\alpha}{1-0.3} \right)^{1/2} \text{ AU} \quad (1)$$

where  $S_{\text{stl}}(\tau_0)$  is the stellar constant at a stellar age  $\tau_0$  when solidification starts.  $S_{\odot}$  is the current solar constant and  $\alpha$  is the planetary albedo.  $a_{\text{cr}}$ , which is about 0.76 in our model, can range from about 0.6 to 0.8 AU, as a result of parameter uncertainties and the atmospheric model used.

Earth is located at a distance of 1 AU from the Sun, which is sufficiently far for it to be a type I planet. Its deep interior would probably have solidified within about 5 Myr, even if it had acquired as much or more water than the current total inventory ( $\sim 1.25\text{--}5M_{\text{EO}}$ )<sup>13</sup>. This would support the rapid solidification and ocean formation condition suggested from geochemical data<sup>14,15</sup>. The overall water budget would have been only slightly affected by hydrodynamic escape, but this modest loss might still have contributed to oxidization of the interior through selective escape of hydrogen over oxygen. If a simple assumption is made

that all of the dissociated oxygen atoms were consumed by oxidation of ferrous iron, the bulk  $\text{Fe}_2\text{O}_3$  content in the early mantle would have increased by up to about 0.07 wt% during the first several million years of Earth's history. This value is comparable to the current  $\text{Fe}_2\text{O}_3$  content in the upper mantle ( $>0.2$  wt%)<sup>16</sup> and is consistent with the idea of an oxidized mantle possibly as early as 4.4 Gyr ago (ref. 17).

Because Venus's orbital distance of about 0.72 AU places it on the border zone, it is not exactly clear which planetary type it should be classified as. However, the significant water loss for a type II planet during the magma-ocean period might explain its dry interior and the fate of the remnant oxygen. The absence of water on its surface<sup>1</sup> and interior<sup>2,3</sup> is conventionally explained in terms of an inherently dry origin<sup>18</sup> or hydrodynamic escape after ocean evaporation<sup>19</sup>. The latter situation corresponds to the case in which Venus is a type I planet in our classification. In this case, the dryness of its interior depends on degassing efficiency during the subsequent evolution. Alternatively, if Venus belongs to type II, the occurrence of planetary desiccation during solidification can explain the dryness of its current mantle<sup>2,3</sup>, even if it accreted with a certain amount of water. The presence of a dry mantle during the earliest stages might also have led to a subsequent tectonic evolution different from that of Earth<sup>20,21</sup> and possibly to the absence of a magnetic field<sup>22</sup>.

One problem associated with the hydrodynamic escape of water is the accumulation of dissociated oxygen in the atmosphere as a result of the preferential escape of hydrogen<sup>23</sup>, which would possibly lead to a shutdown of the hydrogen escape process itself. However, if Venus belongs to type II, the remnant oxygen issue can be avoided, regardless of the initial amount of water, because the abundant surface magma can act as a massive oxygen sink. Although the effects of other gaseous species such as carbon dioxide will need to be addressed in future work, the slight difference in the orbital distances of Earth and Venus might have been the critical factor making these planets totally different after solidification.

Although we have so far discussed only Earth-sized planets orbiting a Sun-like star, the conclusions can be qualitatively extended to all terrestrial planets that experienced global melting through giant impact. More than 750 exoplanets have already been confirmed<sup>24</sup>, and 2,300 candidates<sup>25</sup> have been identified. Future space missions such as the Terrestrial Planet Finder and Darwin will provide us with a catalogue of atmospheres and surface environments for terrestrial planets of various ages. The present results indicate that for habitable planets, rapid ocean formation would have occurred within several million years of planet formation. In addition, the prediction that type II planets undergo a long-lived ( $\sim 100$  Myr) magma-ocean stage suggests that, apart from permanently molten planets very close to the host star, there may be a chance of detecting still-molten terrestrial planets as a consequence of planetary formation, especially around young stars. Although the molten surface would be hard to detect through the thick atmosphere, it can in principle be identified by detecting gaseous species such as NaOH and KOH that are indicative of hot atmospheres<sup>26</sup>.

## METHODS SUMMARY

**Water reservoirs and budgets.** Partition of water between planetary reservoirs was treated in basically the same way as in ref. 6. We used water solubility in basaltic-composition melts calculated from a solubility model in ref. 27. Water partition coefficient into cumulates and mass fraction of interstitially trapped melts were fixed at  $10^{-4}$  and 0.01, respectively.

The water-loss rate was calculated from energy-limited escape rate of hydrogen<sup>12</sup> with a heating efficiency of 0.1 by assuming that all the dissociated oxygen would have been absorbed into the magma ocean. For the effects of the higher heating efficiency, see the online-only Methods and Supplementary Information section 4. We adopted an observational compilation of Sun-like stars for time-dependent extreme ultraviolet radiation<sup>28</sup>.

**Thermal evolution model.** Assuming that the atmosphere has a negligible heat capacity, the thermal evolution model was formulated on the basis of heat balance:

$$C_{p,pl}(T_m) \frac{dT_m}{dt} = -4\pi R^2 (F_{pl} - F_{stl}) \quad (2)$$

where  $C_{p,pl}(T_m)$  is the heat capacity of the planet at potential temperature  $T_m$  of the interior, and  $R$  is the planetary radius.  $F_{pl}$  and  $F_{stl}$  are the outgoing planetary radiation and the incoming net stellar radiation, respectively. Calculation starts from temperature of 3,000 K, which corresponds to a mostly molten state of the mantle, and stops when it reaches the solidus temperature at the surface ( $\sim 1,370$  K). We assumed that the whole silicate portion could transfer heat by convection and therefore contribute to the heat capacity. We obtained the heat capacity, including latent heat of fusion, as a function of potential temperature, from its adiabatic temperature profile calculated according to ref. 29.

The planetary radiation was calculated with a modified radiative–convective equilibrium model of grey atmosphere<sup>11</sup>. We also performed non-grey calculations (Supplementary Information). We adopted a solar standard model for luminosity enhancement with stellar age<sup>30</sup>. A planetary albedo of 0.3 was assumed for the net stellar radiation.

**Full Methods** and any associated references are available in the online version of the paper.

**Received 7 October 2012; accepted 5 April 2013.**

- Lewis, J. S. & Grinspoon, D. H. Vertical distribution of water in the atmosphere of Venus: a simple thermochemical explanation. *Science* **249**, 1273–1275 (1990).
- Nimmo, F. & McKenzie, D. Volcanism and tectonics on Venus. *Annu. Rev. Earth Planet. Sci.* **26**, 23–51 (1998).
- Kiefer, W. S., Richards, M. A., Hager, B. H. & Bills, B. G. A dynamic model of Venus's gravity field. *Geophys. Res. Lett.* **13**, 14–17 (1986).
- Chambers, J. E. & Wetherill, G. W. Making the terrestrial planets: *N*-body integrations of planetary embryos in three dimensions. *Icarus* **136**, 304–327 (1998).
- Canup, R. M. Dynamics of lunar formation. *Annu. Rev. Astron. Astrophys.* **42**, 441–475 (2004).
- Elkins-Tanton, L. T. Linked magma ocean solidification and atmospheric growth for Earth and Mars. *Earth Planet. Sci. Lett.* **271**, 181–191 (2008).
- Matsui, T. & Abe, Y. Evolution of an impact-induced atmosphere and magma ocean on the accreting Earth. *Nature* **319**, 303–305 (1986).
- Zahnle, K. J., Kasting, J. F. & Pollack, J. B. Evolution of a steam atmosphere during Earth's accretion. *Icarus* **74**, 62–97 (1988).
- Abe, Y. & Matsui, T. Evolution of an impact-generated H<sub>2</sub>O–CO<sub>2</sub> atmosphere and formation of a hot proto-ocean on Earth. *J. Atmos. Sci.* **45**, 3081–3101 (1988).
- Kasting, J. F. Runaway and moist greenhouse atmospheres and the evolution of Earth and Venus. *Icarus* **74**, 472–494 (1988).
- Nakajima, S., Hayashi, Y. & Abe, Y. A study on the 'runaway greenhouse effect' with a one-dimensional radiative–convective equilibrium model. *J. Atmos. Sci.* **49**, 2256–2266 (1992).
- Watson, A. J., Donahue, T. M. & Walker, J. C. G. The dynamics of a rapidly escaping atmosphere: applications to the evolution of Earth and Venus. *Icarus* **48**, 150–166 (1981).
- Hirschmann, M. M. Water, melting, and the deep Earth H<sub>2</sub>O cycle. *Annu. Rev. Earth Planet. Sci.* **34**, 629–653 (2006).
- Caro, G. Early silicate Earth differentiation. *Annu. Rev. Earth Planet. Sci.* **39**, 31–58 (2011).
- Cavosie, A. J., Valley, J. W., Wilde, S. A. & Edinburgh Ion Microprobe Facility. Magmatic  $\delta^{18}\text{O}$  in 4400–3900 Ma detrital zircons: a record of the alteration and recycling of crust in the Early Archean. *Earth Planet. Sci. Lett.* **235**, 663–681 (2005).
- Canil, D. & O'Neill, H. St C. Distribution of ferric iron in some upper-mantle assemblages. *J. Petrol.* **37**, 609–635 (1996).
- Trail, D., Watson, E. B. & Tailby, N. D. The oxidation state of Hadean magmas and implications for early Earth's atmosphere. *Nature* **480**, 79–82 (2011).
- Lewis, J. S. Venus: atmospheric and lithospheric composition. *Earth Planet. Sci. Lett.* **10**, 73–80 (1970).
- Kasting, J. F. & Pollack, J. B. Loss of water from Venus. I. Hydrodynamic escape of hydrogen. *Icarus* **53**, 479–508 (1983).
- Mackwell, S. J., Zimmerman, M. E. & Kohlstedt, D. L. High-temperature deformation of dry diabase with applications to tectonics on Venus. *J. Geophys. Res.* **103**, 975–984 (1998).
- Moresi, L. & Solomatov, V. Mantle convection with a brittle lithosphere: thoughts on the global tectonic styles of the Earth and Venus. *Geophys. J. Int.* **133**, 669–682 (1998).
- Nimmo, F. Why does Venus lack a magnetic field? *Geology* **30**, 987–990 (2002).
- Chassefière, E. Hydrodynamic escape of oxygen from primitive atmospheres: applications to the cases of Venus and Mars. *Icarus* **124**, 537–552 (1996).
- The Extrasolar Planets Encyclopedia. <<http://exoplanet.eu/>> (2012).
- Batalha, N. M. *et al.* Planetary candidates observed by *Kepler*. III. Analysis of the first 16 months of data. *Astrophys. J. Suppl. S.* **204**, 24, <http://dx.doi.org/10.1088/0067-0049/204/2/24> (2013).
- Schaefer, L., Lodders, K. & Fegley B. Vaporization of the Earth: application to exoplanet atmospheres. *Astrophys. J.* **755**, 41, <http://dx.doi.org/10.1088/0004-637X/755/1/41> (2012).
- Papale, P. Modeling of the solubility of a one-component H<sub>2</sub>O or CO<sub>2</sub> fluid in silicate liquids. *Contrib. Mineral. Petrol.* **126**, 237–251 (1997).
- Ribas, I., Guinan, E. F., Güdel, M. & Audard, M. Evolution of the solar activity over time and effects on planetary atmospheres. I. High-energy irradiances (1–1700 Å). *Astrophys. J.* **622**, 680–694 (2005).
- Abe, Y. Thermal and chemical evolution of the terrestrial magma ocean. *Phys. Earth Planet. Inter.* **100**, 27–39 (1997).
- Gough, D. O. Solar interior structure and luminosity variations. *Sol. Phys.* **74**, 21–34 (1981).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank J.F. Kasting for constructive comments on the manuscript. We appreciate proofreading and editing assistance from the GCOE programme. This work was supported by Grants-in-Aid for Scientific Research (KAKENHI) from the Ministry of Education, Culture, Sports, Science and Technology (23103003) and from Japan Society for the Promotion of Science (23340168 and 22740291). This study is a part of the PhD thesis of K.H. submitted to the University of Tokyo.

**Author Contributions** K.H. and Y.A. conceived the initial idea. K.H. constructed the coupled model and performed the calculations. Y.A. contributed to the modelling as well. H.G. provided suggestions on exoplanetary science. All authors discussed the results and implications and commented on the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.H. ([keiko@eps.s.u-tokyo.ac.jp](mailto:keiko@eps.s.u-tokyo.ac.jp)).



## METHODS

**Water reservoirs and water budget.** The partitioning of the total water inventory between planetary reservoirs was treated in basically the same way as in ref. 6. Three water reservoirs were considered: atmosphere, magma ocean (a wholly or partly molten region) and solid mantle. The partial pressure of water was calculated at each time step by assuming a solution equilibrium between the atmosphere and the silicate melts at the surface. The water solubility values used were for basaltic melts taken from ref. 31 and were calculated by using the model reported in ref. 27 for a temperature of 2,000 K (see Supplementary Information for the effect of this degassing assumption).

For the incorporation of water into solidifying silicates, two processes were considered: interstitial trapping of water-enriched melts and water partitioning between silicate melts and cumulates. The mass fraction of interstitially trapped melts was assumed to be 1%, as in ref. 6. Although the actual value for a solidifying magma ocean is highly uncertain, a value of 1% is likely to be an overestimate, taking into account the degree of melting estimated for abyssal peridotites when melt separation begins<sup>32</sup>. Few experimental results have been reported on the partition coefficient of water for minerals such as perovskites that make up the deep mantle. We adopted a water partition coefficient of  $10^{-4}$  in our calculations. However, it should be noted that increasing this to  $10^{-2}$  increases the amount of water in the interior at the time of complete solidification by only a factor of 2. This is because interstitial trapping of water-enriched melts dominates.

Because all surface water is in the vapour phase on a solidifying planet, intense hydrodynamic escape of hydrogen is expected to occur during the magma ocean period. We also considered a reduction in the total amount of water due to hydrodynamic escape. The water loss rate was calculated from the energy-limited escape rate for hydrogen<sup>12</sup> by assuming that all of the oxygen produced by dissociation of water vapour would be lost from the atmosphere by oxidation of the magma ocean. Extreme ultraviolet radiation levels for the host star were based on compiled observational data for Sun-like stars, taken from ref. 28, in which the decay with time is expressed as a power-law function of stellar age. A heating efficiency of 0.1 was used throughout all the calculations. Using the higher heating efficiency has less effect on the evolution of type I planets, whereas the solidification time of type II planets becomes shorter in inverse proportion to the heating efficiency (see Supplementary Information).

**Thermal evolution model.** A high degree of melting implies an extremely turbulent magma ocean with a viscosity probably as low as 0.1 Pa (ref. 33). This suggests that the magma ocean is effective in transporting internal heat to the surface. Planetary radiation into space would therefore be the main factor limiting the heat flux from the deep magma ocean. We assumed that the heat flux is given by the difference between the outgoing planetary radiation  $F_{\text{pl}}$  and incoming net stellar radiation  $F_{\text{stl}}$ . On the assumption that the atmosphere has a negligible heat capacity, a thermal evolution model can be formulated on the basis of the heat balance for a planet,

$$C_{\text{p,pl}}(T_{\text{m}}) \frac{dT_{\text{m}}}{dt} = -4\pi R^2 (F_{\text{pl}} - F_{\text{stl}}) \quad (2)$$

where  $C_{\text{p,pl}}(T_{\text{m}})$  is the heat capacity of a planet with an interior potential temperature  $T_{\text{m}}$ , and  $R$  is the planetary radius. Heat flux from the core and heat production by radiogenic elements are neglected. The simple assumption is made that the potential temperature matches the surface temperature  $T_{\text{s}}$ —that is,  $T_{\text{m}} = T_{\text{s}}$ —during the solidification period of the magma ocean. The calculation starts with potential and surface temperatures of 3,000 K, implying a mostly molten mantle, and stops when the solidus temperature at the surface ( $\sim 1,370$  K) is reached.

We calculated the planetary heat capacity, including the latent heat of fusion, as a function of the potential temperature by assuming that the entire silicate portion can transfer heat by convection, and not only the molten part as considered in ref. 6. The heat capacity calculated in this way corresponds to the maximum value, thus providing a conservative estimate of the maximum solidification time for a type I planet. However, both assumptions about the interior temperature profile result in an almost identical solidification time for a type II planet, because the solidification rate is mainly determined not by the heat capacity but by the water loss rate. The adiabatic thermal structure in the interior was calculated according to ref. 29 by using the parameters therein, except for recent experimental data for the solidus and liquidus temperatures at great depth (see Supplementary Information). Differentiation of the mantle and its mineralogy on solidification were not considered in the calculations of the temperature profile, because these would be expected to make a smaller contribution to the overall heat capacity of the planet. See the next section for calculations of planetary radiation and net stellar radiation.

**Atmospheric model.** Planetary radiation was calculated by using a modified version of the radiative–convective equilibrium model in ref. 11, which was originally designed for a steam atmosphere saturated with water vapour and for moderate surface temperatures. However, in the present study the surface temperatures considered are extremely high, exceeding the critical temperature of water. In addition, the relative humidity of the atmosphere can vary because the partial pressure of water is determined by solution equilibrium at the surface. We have extended the grey atmosphere model in ref. 11 to be applicable to the conditions considered herein. The critical point of water was simply treated as the temperature above which no water condensation would occur. For atmospheric layers that were unsaturated by water vapour or for which the temperature was above the critical point, the temperature lapse rate in the troposphere was assumed to be a dry adiabatic temperature gradient, whereas for saturated parcels the lapse rate was assumed to be a pseudo-moist adiabatic temperature gradient, as in ref. 11. The atmosphere is considered to consist of 1 bar of  $\text{N}_2$  in addition to water vapour, the amount of which is determined by the water budget. See Supplementary Information for results calculated with a non-grey atmospheric model in ref. 9, instead of our simple grey model.

For calculations of the net stellar radiation, the solar standard model<sup>30</sup> was adopted to describe the change in stellar luminosity with time. The planetary albedo was fixed at 0.3 in all calculations. The actual value of planetary albedo is highly uncertain, especially because of the unpredictability of cloud coverage with a one-dimensional atmospheric model. In the previous studies (refs 9, 10) a planetary albedo for a thick steam atmosphere was estimated as large as 0.4–0.5, whereas a recent study<sup>34</sup> with a non-grey atmospheric model using the HITEMP database has shown that including weak  $\text{H}_2\text{O}$  lines significantly lowers the planetary albedo.

31. Presnall, D. C. & Hoover, J. D. in *Magmatic Processes: Physicochemical Principles* (ed. Mysen, B. O.) vol. 1, 75–89 (Geochemical Society Special Publication, 1984).
32. Johnson, K. T. M. & Dick, H. J. B. Open system melting and temporal and spatial variation of peridotite and basalt at the Atlantis II fracture zone. *J. Geophys. Res.* **97**, 9219–9241 (1992).
33. Solomatov, V. in *Evolution of the Earth* (ed. Stevenson, D.) 91–119 (Treatise on Geophysics 9, Elsevier, 2007).
34. Kopparapu, R. K. *et al.* Habitable zones around main-sequence stars: new estimates. *Astrophys. J.* **765**, 131, <http://dx.doi.org/10.1088/0004-637X/765/2/131> (2013).

# Palaeontological evidence for an Oligocene divergence between Old World monkeys and apes

Nancy J. Stevens<sup>1,2</sup>, Erik R. Seiffert<sup>3</sup>, Patrick M. O'Connor<sup>1,2</sup>, Eric M. Roberts<sup>4</sup>, Mark D. Schmitz<sup>5</sup>, Cornelia Krause<sup>1,2</sup>, Eric Gorscak<sup>6</sup>, Sifa Ngasala<sup>7</sup>, Tobin L. Hieronymus<sup>8</sup> & Joseph Temu<sup>9</sup>

Apes and Old World monkeys are prominent components of modern African and Asian ecosystems, yet the earliest phases of their evolutionary history have remained largely undocumented<sup>1</sup>. The absence of crown catarrhine fossils older than ~20 million years (Myr) has stood in stark contrast to molecular divergence estimates of ~25–30 Myr for the split between Cercopithecoidea (Old World monkeys) and Hominoidea (apes), implying long ghost lineages for both clades<sup>2–4</sup>. Here we describe the oldest known fossil ‘ape’, represented by a partial mandible preserving dental features that place it with ‘nyanzapithecine’ stem hominoids. Additionally, we report the oldest stem member of the Old World monkey clade, represented by a lower third molar. Both specimens were recovered from a precisely dated 25.2-Myr-old stratum in the Rukwa Rift, a segment of the western branch of the East African Rift in Tanzania. These finds extend the fossil record of apes and Old World monkeys well into the Oligocene epoch of Africa, suggesting a possible link between diversification of crown catarrhines and changes in the African landscape brought about by previously unrecognized tectonic activity<sup>5</sup> in the East African rift system.

The late Oligocene represents the least-sampled temporal interval in primate evolutionary history<sup>3</sup>, with only a handful of primates described from all of Afro-Arabia<sup>6–10</sup>. Possible reasons for an end-Palaeogene gap in the fossil record include limited deposits of appropriate age, particularly from Africa below the equator, complicated by densely vegetated topography in more tropical environments. As a result of this sampling bias, detailed understanding of the early diversification of Old World monkeys and apes has remained elusive. In particular, fossils from this interval are critical for testing the hypothesis of a late Palaeogene (~25–30 Myr) hominoid–cercopithecoid divergence, a result repeatedly retrieved by molecular studies<sup>2–4</sup>. Recent discoveries from the Rukwa Rift Basin in southwestern Tanzania provide critical data for testing these ideas by revealing a novel glimpse into late Oligocene terrestrial ecosystems from Africa below the equator<sup>10–12</sup>.

The Rukwa Rift Basin (Fig. 1) records one of the thickest accumulations of sedimentary rock in the entire East African rift system (EARS)<sup>11</sup>. Work over the past decade has documented a complex and long-lived history of the western branch of the EARS<sup>5,11</sup> containing continental faunas of both Cretaceous<sup>12</sup> and Oligocene<sup>10</sup> age, the latter spanning ~24–26 Myr (see Supplementary Information, section 1). Palaeontological field research in 2011–12 resulted in the recovery of two well-preserved primate fossils from the Nsungwe 2B locality. These discoveries provide critical data for resolving disparities between molecularly derived divergence estimates and the primate fossil record.

Primates Linnaeus, 1758  
Anthropoidea Mivart, 1864  
Catarrhini Geoffroy, 1812  
Cercopithecoidea Gray, 1821  
*Nsungwepithecus* gen. nov.

**Etymology.** Prefix ‘Nsungwe’ in reference to the name of the geological formation from which the specimen was recovered; ‘pithecus’ a common primate suffix derived from the Greek *pithekos* (ape).

**Diagnosis.** Differs from all other fossil cercopithecoids in exhibiting the following combination of features on the lower third molar (M<sub>3</sub>): larger than all known ‘victoriapithecids’ other than *Zaltanpithecus*<sup>13,14</sup>; unbifurcated mesial root; low rounded cusps with pronounced buccal flare; hypoconulid pronounced and centrally positioned; deeply incised distal buccal cleft that extends to the crown base; incomplete bilophodonty, with a small notch in the lophid connecting the protoconid and metaconid and absence of a hypolophid connecting the entoconid and hypoconid; lower degree of mesiodistal elongation and basal inflation than in *Noropithecus*<sup>14</sup>; marked buccal enamel wrinkling extending onto the median buccal ridge; shallow and crenulated lingual notch; and a proliferation of accessory cusplids along the postmetacristid and around the entoconid, including three cusplids situated in the talonid basin.

*Nsungwepithecus gunnelli* sp. nov.

**Etymology.** Specific name is in honour of Gregg F. Gunnell for his many contributions to primate palaeontology.

**Holotype.** RRBP (Rukwa Rift Basin Project) 11178, left partial mandible preserving M<sub>3</sub> (Fig. 2b; see also Supplementary Information, section 2).

**Locality and horizon.** Oligocene Nsungwe Formation, locality Nsungwe 2B, near the town of Mbeya, southwestern Tanzania (Fig. 1a). The site is situated 30 m above the contact between the Utengule and Songwe members of the Nsungwe Formation. The age of the fossil-bearing unit is tightly constrained between two volcanic tuffs dated by U-Pb CA-TIMS (U-Pb chemical abrasion thermal ionization mass spectrometry) geochronology at 25.237 and 25.214 Myr ago (Fig. 1c and Supplementary Information).

**Diagnosis.** As for genus. For additional description and metrics, see Supplementary Information.

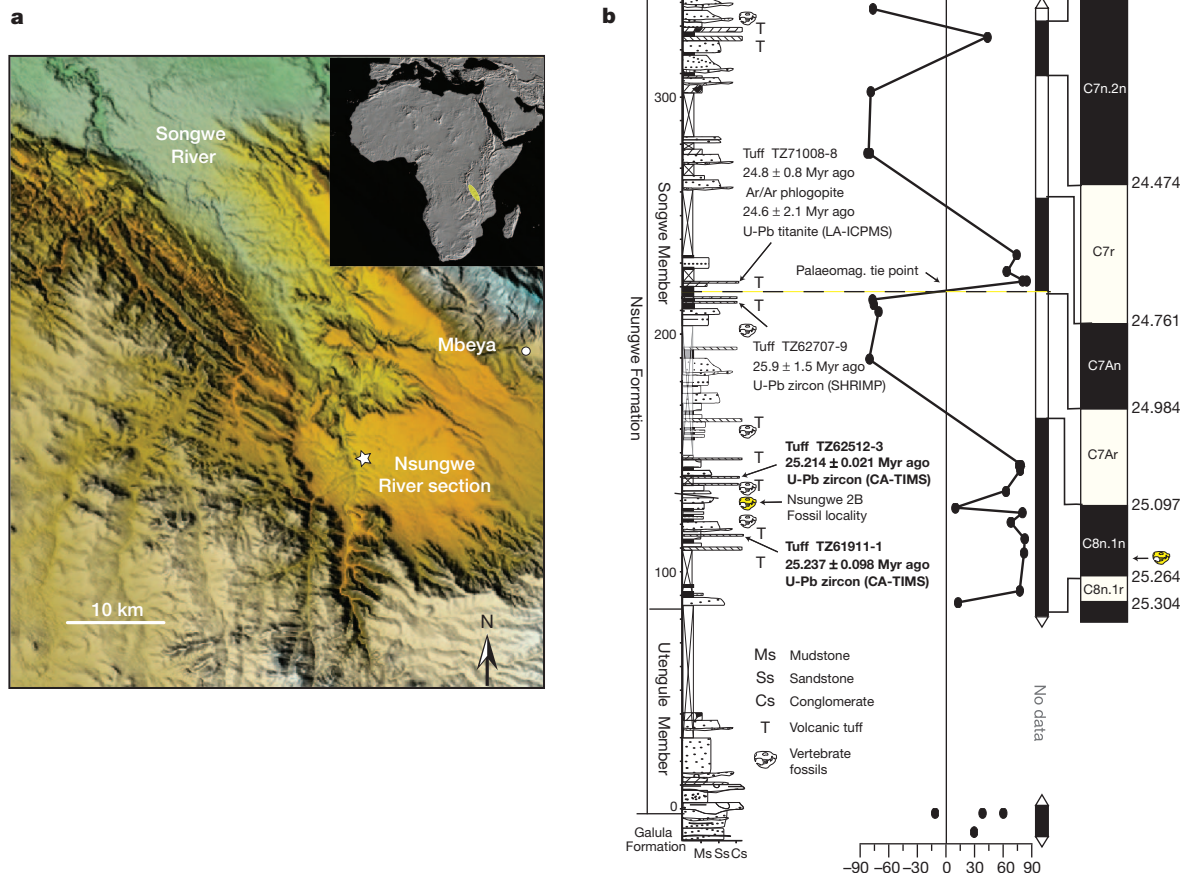
Hominoidea Gray, 1825

*Rukwapithecus* gen. nov.

**Etymology.** Prefix ‘Rukwa’ in reference to the Rukwa Rift Basin from which the specimen was recovered; ‘pithecus’ a common primate suffix derived from the Greek *pithekos* (ape).

**Diagnosis.** Differs from all other extinct catarrhines in the following combination of characters: long, high-crowned and obliquely implanted lower fourth premolar, with mesial basin elevated high above the longer and more distolingually oriented talonid basin; mesiodistally elongate and crenulated lower molars that increase in length and width distally, and that are subrectangular and waisted in outline; mesial position of lower molar protoconid and hypoconid relative to lingual cusps; deep lower molar hypoflexid formed in part by a pronounced buccal cingulid that surrounds the protoconid mesially, but blends onto the buccal surface of the hypoconid; small lower molar metastylid (= mesoconid<sup>15</sup>) distolingual to the metaconid that is more pronounced on M<sub>2</sub>–M<sub>3</sub>;

<sup>1</sup>Department of Biomedical Sciences, Heritage College of Osteopathic Medicine, Ohio University, Athens, Ohio 45701, USA. <sup>2</sup>Center for Ecology and Evolutionary Studies, Ohio University, Athens, Ohio 45701, USA. <sup>3</sup>Department of Anatomical Sciences, Stony Brook University, Stony Brook, New York 11794, USA. <sup>4</sup>School of Earth and Environmental Sciences, James Cook University, Townsville, Queensland 4811, Australia. <sup>5</sup>School of Geosciences, Boise State University, Boise, Idaho 83725-1535, USA. <sup>6</sup>Department of Biological Sciences, Ohio University, Athens, Ohio 45701, USA. <sup>7</sup>Department of Geological Sciences, Michigan State University, East Lansing, Michigan 48824, USA. <sup>8</sup>Department of Anatomy and Neurobiology, Northeast Ohio Medical University, Rootstown, Ohio 44272, USA. <sup>9</sup>Tanzania Antiquities Unit, PO Box 2280, Dar es Salaam, Tanzania.



**Figure 1 | Location and stratigraphy of the primate-bearing locality (Nsungwe 2B), in southwestern Tanzania.** **a**, Digital elevation model of the study area based on Shuttle Radar Topography Mission data (SRTM). Inset map highlights the position of the Rukwa Rift Basin in eastern Africa (yellow oval). **b**, Measured stratigraphic section through the Nsungwe Formation, showing the position of Nsungwe 2B in yellow with the positions of two

recently dated (via U-Pb CA-TIMS; bold type) carbonatite tuffs and several other dated tuffs (left), interpreted palaeomagnetic reversal stratigraphy of ref. 5 (virtual geomagnetic pole latitude (VGP lat.)) (centre), and ages (Myr ago; right) derived from the global polarity timescale (GPTS). Black bars, normal polarity; white bars, reversed polarity (see Supplementary Information for additional geological details).

accessory cuspules in lingual notches of  $M_2$  and  $M_3$ ; deep lingual notch and distolingual fovea on  $M_1$ – $M_3$ ; large, well-individuated and buccally positioned hypoconulid on  $M_1$ – $M_3$ ; cresting between entoconid and hypoconulid weak or absent on  $M_1$ – $M_3$ ; and  $M_3$  massive and highly crenulated, slightly tapering to a broad and rounded distal margin.

*Rukwapithecus fleaglei* sp. nov.

**Etymology.** Specific name is in honour of John G. Fleagle, for his many contributions to the study of primate morphology, behaviour and evolution.

**Holotype.** RRBP 12444A, a right mandible bearing lower fourth premolar ( $P_4$ ) through to  $M_3$  and the ascending ramus (Fig. 2i and Supplementary Information, section 3).

**Locality and horizon.** Oligocene Nsungwe Formation, locality Nsungwe 2B (as described above).

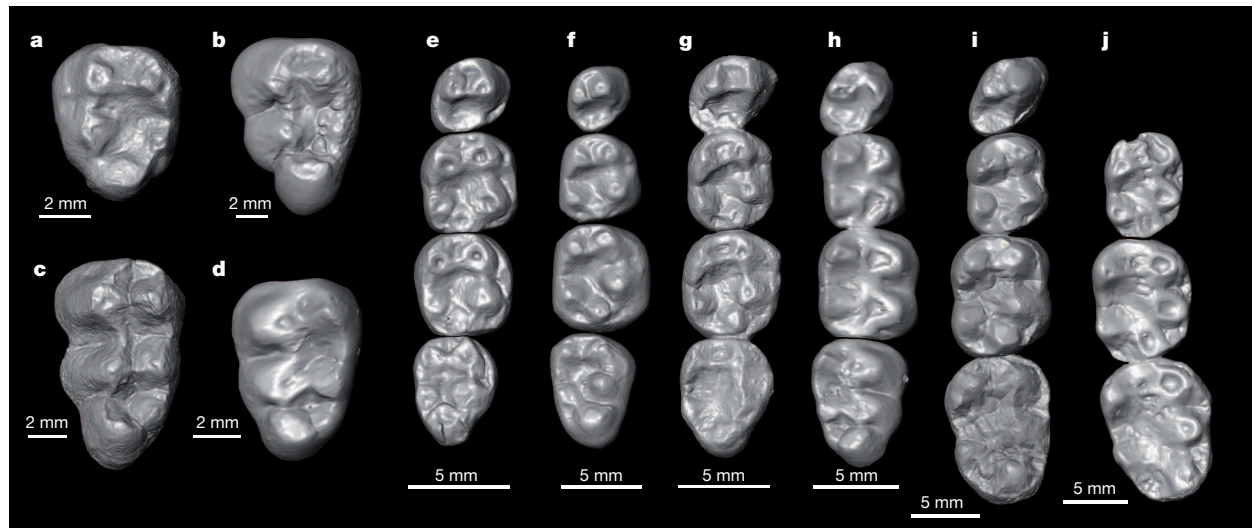
**Diagnosis.** As for genus. For additional description and metrics, see Supplementary Information.

Morphological features defining the earliest crown catarrhines have largely been a matter of speculation, with victoriapithecids and proconsuloids often defined primarily by the shared retention of primitive characters relative to later forms, rather than by demonstrable synapomorphies<sup>14–17</sup>. Poor resolution of the phylogenetic branching pattern among fossil forms near the base of the cercopithecoid–hominoid split is further complicated by an unbalanced fossil record for the two

groups, with early cercopithecoids represented by relatively few early-mid Miocene taxa<sup>14</sup> and Miocene apes exhibiting greater taxonomic diversity<sup>15</sup>. The presence of taxa as distinctive as *Prohylobates*, *Proconsul* and *Rangwapithecus* in the African early Miocene indicates that the cercopithecoid–hominoid diversification initiated during the Oligocene<sup>15,18</sup>, yet the only previously described late Oligocene catarrhines, *Kamoyapithecus* and *Saadanius*, are generally regarded as stem forms rather than members of the catarrhine crown clade<sup>6,8,17</sup>. As such, *Rukwapithecus* and *Nsungwepithecus* are the first described primates that document the presence of crown catarrhines as early as ~25 Myr ago.

*Nsungwepithecus* represents the first cercopithecoid old enough to confirm the late Palaeogene crown catarrhine divergence estimates derived from molecular studies<sup>2–4</sup>. *Nsungwepithecus* shares with victoriapithecids numerous features of lower molar morphology including deeply incised buccal clefts, a high degree of buccal flare, and the lack of a buccal cingulid. Before the late Miocene, the published cercopithecoid record has largely been limited to rare and incomplete materials of *Prohylobates* and *Zaltanpithecus* collected from the early-middle Miocene of northern Africa<sup>13,14,18–20</sup>, together with a spectacular array of over 2,500 specimens from a single taxon (*Victoriapithecus*) from mid-Miocene deposits in eastern Africa<sup>14,16,19,21</sup>. Additional cercopithecoid diversity has recently been recognized in early-middle Miocene faunas from Kenya and Uganda, represented by as many as



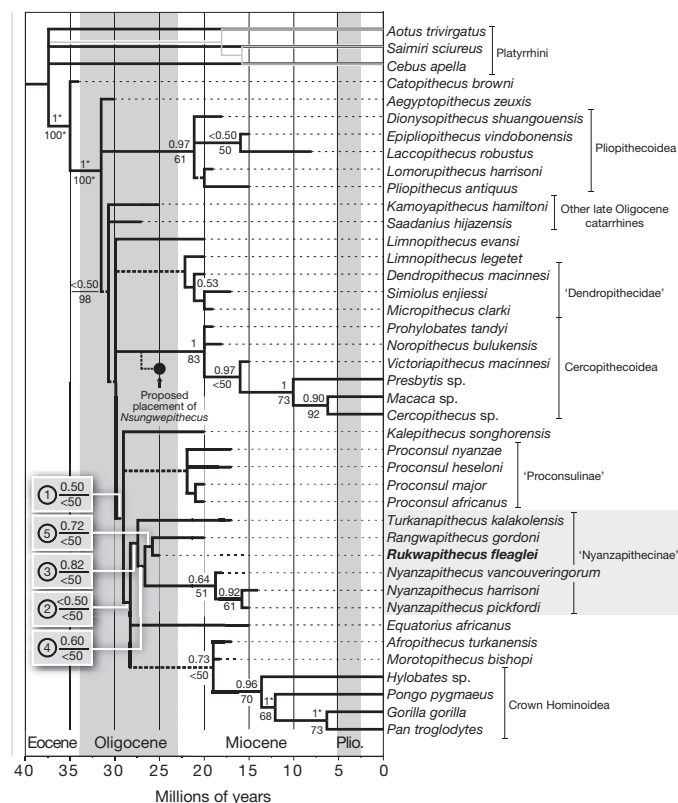


**Figure 2 | Comparison of Nsungwe Formation primates with representative stem and crown catarrhines.** **a**, *Propliopithecus* sp. (TQ 4, early Oligocene of Oman), right M<sub>3</sub>, reversed for comparison; **b**, *Nsungwepithecus gunnelli* gen. et sp. nov. (RRBP 11178, late Oligocene of Tanzania), left M<sub>3</sub>; **c**, *Noropithecus bulukensis* (KNM-WS 12642, early Miocene of Kenya), right M<sub>3</sub>, reversed for comparison; **d**, *Victoriapithecus macinnesi* (KNM-MB 18993, middle Miocene of Kenya), right M<sub>3</sub>, reversed for comparison; **e**, *Propliopithecus haeckeli* (SMN 12638, early Oligocene (?) of Egypt), right P<sub>4</sub>-M<sub>3</sub>; **f**, *Aegyptopithecus zeuxis*

(DPC 3056, early Oligocene of Egypt), right P<sub>4</sub>-M<sub>3</sub>; **g**, *Kalepithecus songhorensis* holotype (KNM-SO 378, early Miocene of Kenya), right P<sub>4</sub>-M<sub>3</sub>; **h**, *Victoriapithecus macinnesi* (KNM-MB 18993, middle Miocene of Kenya), right M<sub>1</sub>-M<sub>3</sub> and reversed left P<sub>4</sub>; **i**, *Rukwapithecus fleaglei* gen. et sp. nov. (RRBP 12444A, late Oligocene of Tanzania), right P<sub>4</sub>-M<sub>3</sub>; **j**, *Rangwapithecus gordonii* (KNM-SO 463, early Miocene of Kenya), right M<sub>1</sub>-M<sub>3</sub>. See Supplementary Information section 4 for imaging protocols, and Supplementary Videos 1 and 2 for additional views of Nsungwe specimens.

three species within the bilophodont genus *Noropithecus*<sup>14</sup>. Collectively, these fossils have been grouped in the Victoriapithecidae (a basal cercopithecoidean group), with the acknowledgement that additional, more complete, fossils may reveal this to be a paraphyletic assemblage<sup>14</sup>. The recovery of a stem cercopithecoidean older than 25 Myr ago significantly extends the record of this clade, documenting the presence of a relatively large and incompletely bilophodont monkey in the Oligocene of eastern Africa.

*Rukwapithecus* shares two features with the Miocene apes and extant hominoids that are not present in cercopithecoideans or Fayum stem catarrhines<sup>15,16</sup>: buccal position of the M<sub>2</sub> hypoconulid, and mesial migration of cusps on the buccal side of lower molars such that the hypoconid is positioned opposite the lingual notch between the metaconid and the entoconid. In particular, *Rukwapithecus* shares numerous features with the early Miocene *Rangwapithecus*<sup>15,17,22,23</sup>, including cusp position and wear pattern, degree of crenulation and cingulid development, oblique orientation of the cristid obliqua on M<sub>1</sub> and M<sub>2</sub>, deep hypoflexid, deep distolingual fovea, and an enlarged M<sub>3</sub>.



**Figure 3 | Phylogenetic placement of *Rukwapithecus*, new genus.**

Temporally calibrated Adams consensus of four equally parsimonious trees recovered in PAUP 4.0b10 and the 'allcompat' tree calculated in MrBayes 3.2, based on analysis of the modified and taxonomically expanded 191-character morphological character matrix of ref. 26 (see Supplementary Information for details; thick dashed lines indicate branches that are not present in the strict consensus of all five trees; asterisks next to support values indicate that that node was constrained in either the parsimony or Bayesian analysis). Eocene *Catopithecus* was constrained as the most basal stem catarrhine due to its retention of several plesiomorphies that demonstrably evolved convergently in later catarrhines and the extant platyrrhine outgroups<sup>27</sup>. Both analyses congruently placed *Rukwapithecus* as a nyanzapithecine stem hominoid, and, within 'Nyanzapithecinae', as the sister taxon of early Miocene *Rangwapithecus*. We obtained the same placement of *Rukwapithecus* when *Lomropithecus* was removed and scorings for *Afropithecus* and *Morotopithecus* were combined into a single operational taxonomic unit, following an alternative taxonomic hypothesis of ref. 15. Numbers above and below branches are Bayesian posterior probability values and bootstrap values, respectively. We place quotation marks around 'Nyanzapithecinae' because reports of more complete materials of *Mabokopithecus* may eventually necessitate reassignment of some or all *Nyanzapithecus* species to the former genus<sup>15,17</sup> due to taxonomic priority<sup>19</sup>, and may result in a name change for 'Nyanzapithecinae'. The new genus *Nsungwepithecus* was not included in the phylogenetic analyses, and is grafted onto the tree in its proposed placement as the most basal known stem cercopithecoidean. Unambiguous synapomorphies for nodes numbered 1–5 are provided in the Supplementary Information. Divergence dates within crown Cercopithecoidea and crown Hominoidea are based on the molecular dating analysis of ref. 4 with independent rates and soft bounds.

Such features suggest that the lower molar pattern in the *Rukwapithecus*–*Rangwapithecus* clade was fairly conserved across the Oligocene–Miocene transition. *Rukwapithecus* nevertheless differs from *Rangwapithecus* in a number of features, exhibiting for example a narrower mesial fovea, a less tapered distal margin of  $M_3$ , and relatively weak cresting between the lower molar entoconid and hypoconulid. Indeed, bootstrap support for the *Rangwapithecus*–*Rukwapithecus* clade in our parsimony analyses is very low (<50%), arguing against a particularly close (genus level) relationship (Fig. 3 and Supplementary Information, sections 5–7). Parsimony and Bayesian phylogenetic analyses<sup>24–27</sup> place *Rukwapithecus* as a stem hominoid nested within the ‘nyanzapithecine’ clade (sensu Harrison<sup>15</sup>, see Fig. 3), but this result is not particularly robust (see Fig. 3 support values and Supplementary Information). In light of this, we cautiously place *Rukwapithecus* in both ‘Nyanzapithecinae’ and Hominoidea, but recognize that additional data from other parts of the dentition, cranium and postcranium are necessary for further testing and refining these hypotheses.

The Cenozoic era of Africa records a remarkable and deep record of environmental change. During this time, tectonic activity in the prominent East African rift system<sup>5</sup>, uplift of the African plateau<sup>5,28</sup>, and climate aridification<sup>28</sup> had profound implications for Africa’s resident biota. Near the Oligocene–Miocene boundary, collision between the Afro-Arabian and Eurasian landmasses initiated periodic faunal interchange that contributed to the eventual replacement of many resident forms by immigrant species<sup>29</sup>. Given the paucity of palaeontological data from the 22–30-Myr interval in Afro-Arabia, fossils from the Rukwa Rift provide a rare window into Palaeogene catarrhine diversity during this period of dramatic change in African terrestrial ecosystems, with *Nsungwepithecus* and *Rukwapithecus* together comprising 40% of described late Oligocene anthropoid taxa. The precisely dated stratigraphy of the site suggests that early hominoid and cercopithecoid evolution in eastern Africa took place against the backdrop of previously unrecognized tectonic uplift in the western branch of the EARS<sup>5</sup>, coinciding with the global late Oligocene warming event<sup>30</sup>, and pre-dating larger-scale faunal shifts that intensified later in the Miocene.

## METHODS SUMMARY

Rukwa specimens (RRBP 12444A and RRBP 11178) were scanned at the Ohio University MicroCT (OU $\mu$ CT) facility in Athens, Ohio, using a GE eXplore Locus *in vivo* small animal MicroCT scanner. The *Rukwapithecus* type specimen (RRBP 12444A) was scanned at a slice thickness of 90  $\mu$ m, 80 kV, 495 mA yielding a voxel size of  $0.09 \times 0.09 \times 0.09$  mm. For a more detailed reconstruction of occlusal surfaces RRBP 12444A was also scanned at a slice thickness of 20  $\mu$ m, 80 kV, 495 mA. The latter protocol was also used for the *Nsungwepithecus* type specimen (RRBP 11178), yielding a voxel size of  $0.02 \times 0.02 \times 0.02$  mm for high-resolution scans. The resulting volume data (in VFF-format) were exported from MicroView 2.2 (open-source software developed by GE; <http://www.sourceforge.net>) and imported into Avizo 6.3 (Visualization Sciences Group) for image segmentation, visualization and manipulation. Protocols for phylogenetic analysis and high-precision CA-TIMS U–Pb zircon ages for the fossiliferous locality are provided in Supplementary Information.

Received 21 February; accepted 5 April 2013.

Published online 15 May 2013.

- Jablonski, N. & Frost, S. Cercopithecoidea. In *Cenozoic Mammals of Africa* (eds Werdelin, L. & Sanders, W. J.) 393–428 (Univ. California Press, 2010).
- Steiper, M. E. & Young, N. M. Timing primate evolution: Lessons from the discordance between molecular and paleontological estimates. *Evol. Anthropol.* **17**, 179–188 (2008).
- Wilkinson, R. D. *et al.* Dating primate divergences through an integrated analysis of paleontological and molecular data. *Syst. Biol.* **60**, 16–31 (2011).
- Springer, M. S. *et al.* Macroevolutionary dynamics and historical biogeography of primate diversification inferred from a species supermatrix. *PLoS ONE* **7**, e49521 (2012).
- Roberts, E. M. *et al.* Initiation of the western branch of the East African Rift coeval with the eastern branch. *Nature Geosci.* **5**, 289–294 (2012).
- Leakey, M. G., Ungar, P. S. & Walker, A. A new genus of large primate from the Late Oligocene of Lothidok, Turkana District, Kenya. *J. Hum. Evol.* **28**, 519–531 (1995).

- Stevens, N. J. *et al.* An anthropoid primate humerus from the Rukwa Rift Basin, Paleogene of southwestern Tanzania. *J. Vert. Paleontol.* **25**, 986–989 (2005).
- Zalmout, I. S. *et al.* New Oligocene primate from Saudi Arabia and the divergence of apes and Old World monkeys. *Nature* **466**, 360–364 (2010).
- Ducrocq, S., Manthi, F. K. & Lihoreau, F. First record of a parapithecoid primate from the Oligocene of Kenya. *J. Hum. Evol.* **61**, 327–331 (2011).
- Stevens, N. J. *et al.* In *Elwyn L. Simons: A Search for Origins* (eds Fleagle, J. G. & Gilbert, C. C.) 159–180 (Springer, 2008).
- Roberts, E. M. *et al.* Sedimentology and depositional environments of the Red Sandstone Group, Rukwa Rift Basin, Tanzania: New insight into Cretaceous and Paleogene terrestrial ecosystems and tectonics in sub-Equatorial Africa. *J. Afr. Earth Sci.* **57**, 179–212 (2010).
- O’Connor, P. M. *et al.* The evolution of mammal-like crocodyliiforms in the Cretaceous of Gondwana. *Nature* **466**, 748–751 (2010).
- Benefit, B. R. The biostratigraphy and paleontology of fossil cercopithecoids from eastern Libya. *Geol. East Libya* **3**, 247–266 (2008).
- Miller, E. R. *et al.* Systematics of early and middle Miocene Old World monkeys. *J. Hum. Evol.* **57**, 195–211 (2009).
- Harrison, T. Dendropithecoidea, Proconsuloidea and Hominoidea. In *Cenozoic Mammals of Africa* (eds Werdelin, L. & Sanders, W. J.) 429–469 (Univ. California Press, 2010).
- Benefit, B. R. *Victoriapithecus*: The key to Old World monkey and catarrhine origins. *Evol. Anthropol.* **7**, 155–174 (1999).
- Harrison, T. In *The Primate Fossil Record* (ed. Hartwig, W. C.) 311–338 (Cambridge Univ. Press, 2002).
- Fourtau, R. Contribution à l’étude des vertébrés Miocène de l’Égypte (Survey Dept. Min. Finance Govt. Press, 1918).
- Von Koenigswald, G. H. R. In *Fossil Vertebrates of Africa* (ed. Leakey, L. S. B.) 39–52 (Academic, 1969).
- Delson, E. *Prohylobates* (Primates) from the Early Miocene of Libya: A new species and its implications for cercopithecoid origins. *Geobios* **12**, 725–733 (1979).
- Benefit, B. R. The permanent dentition and phylogenetic position of *Victoriapithecus* from Maboko Island, Kenya. *J. Hum. Evol.* **25**, 83–172 (1993).
- Andrews, P. New species of *Dryopithecus* from Kenya. *Nature* **249**, 188–190 (1974).
- Nengo, I. O. & Rae, T. C. New hominoid fossils from the early Miocene site of Songhor, Kenya. *J. Hum. Evol.* **23**, 423–429 (1992).
- Swofford, D. L. *PAUP\* Phylogenetic Analysis Using Parsimony (\*and Other Methods)*, Version 4 (Sinauer Associates, 1998).
- Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574 (2003).
- Rossie, J. B. & MacLatchy, L. A new pliopithecoid genus from the early Miocene of Uganda. *J. Hum. Evol.* **50**, 568–586 (2006).
- Seiffert, E. R. Early primate evolution in Afro-Arabia. *Evol. Anthropol.* **21**, 239–253 (2012).
- Sepulchre, P. *et al.* Tectonic uplift and eastern African aridification. *Science* **313**, 1419–1423 (2006).
- Leakey, M., Grossman, A., Gutiérrez, M. & Fleagle, J. G. Faunal change in the Turkana Basin during the late Oligocene and Miocene. *Evol. Anthropol.* **20**, 238–253 (2011).
- Zachos, J. *et al.* Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* **292**, 686–693 (2001).

Supplementary Information is available in the online version of the paper.

**Acknowledgements** We thank D. Kamamba, F. Ndunguru, J. Temba, P. Msemwa, I. Marobhe, E. Mbende and the Tanzania Commission for Science and Technology for support; E. Lund for assisting with specimen preparation; S. Egberts, R. Felice, E. Lund, H. O’Brien and J. Sidote for field assistance in 2011 and/or 2012; M. Gottfried for earlier project contributions; H. Fässler, T. Plattner, A. Njao and S. and T. Greenaway for their support; M. Cortese-Hering, M. Dawson, J. Eastman, J. Fleagle, G. Gunnell, J. A. Holman, D. Krause, S. Howard, L. Jolley, D. Miles, L. Robbins, N. Sauer, C. Seiffert, E. Simons and P. Wright for helpful discussions; G. Gunnell, F. Manthi, E. Mbua and M. Muungu for specimen access; E. Delson for *Victoriapithecus*  $\mu$ CT data; L. Halenar, W. Holloway and S. Maiolino for  $\mu$ CT assistance; J. Sattler for specimen photography; and M. Antón for scientific artwork. D. DeBlieux and V. Simons discovered the Nsungwe 2 locality in 2007; R. Felice co-discovered RRBP 11178. Research was supported by US National Science Foundation (EAR-0617561, EAR/IF-0933619, BCS-1127164), National Geographic Society (CRE), Louis B. Leakey Foundation, Ohio University Heritage College of Osteopathic Medicine, and Ohio University Office of Research and Sponsored Programs.

**Author Contributions** N.J.S., P.M.O. and E.M.R. developed the Nsungwe primate project of the RRBP. N.J.S., E.M.R., P.M.O., E.G., T.L.H., C.K., S.N. and J.T. conducted excavations. N.J.S., P.M.O., E.R.S., E.M.R., M.D.S. and C.K. conducted the research. N.J.S., E.R.S., P.M.O., M.D.S. and E.M.R. wrote the manuscript and Supplementary Information.

**Author Information** The ZooBank numbers for each taxon are as follows: urn:lsid:zoobank.org:act:15FF7EAO-2EAE-4FDO-A514-C7624FCE66CA (*Nsungwepithecus gunnelli*) and urn:lsid:zoobank.org:act:4A7B64DF-55D6-4F77-ACB3-82F01F993E7C (*Rukwapithecus fleaglei*). Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.J.S. (stevensn@ohio.edu).

# Long-term warming restructures Arctic tundra without changing net soil carbon storage

Seeta A. Sistla<sup>1</sup>, John C. Moore<sup>2,3</sup>, Rodney T. Simpson<sup>2</sup>, Laura Gough<sup>4</sup>, Gaius R. Shaver<sup>5</sup> & Joshua P. Schimel<sup>1</sup>

High latitudes contain nearly half of global soil carbon, prompting interest in understanding how the Arctic terrestrial carbon balance will respond to rising temperatures<sup>1,2</sup>. Low temperatures suppress the activity of soil biota, retarding decomposition and nitrogen release, which limits plant and microbial growth<sup>3</sup>. Warming initially accelerates decomposition<sup>4–6</sup>, increasing nitrogen availability, productivity and woody-plant dominance<sup>3,7</sup>. However, these responses may be transitory, because coupled abiotic–biotic feedback loops that alter soil-temperature dynamics and change the structure and activity of soil communities, can develop<sup>8,9</sup>. Here we report the results of a two-decade summer warming experiment in an Alaskan tundra ecosystem. Warming increased plant biomass and woody dominance, indirectly increased winter soil temperature, homogenized the soil trophic structure across horizons and suppressed surface-soil-decomposer activity, but did not change total soil carbon or nitrogen stocks, thereby increasing net ecosystem carbon storage. Notably, the strongest effects were in the mineral horizon, where warming increased decomposer activity and carbon stock: a ‘biotic awakening’ at depth.

The carbon balance of northern ecosystems will be regulated ultimately by the difference between carbon gains through greater primary productivity and losses through increased soil organic matter (SOM) decomposition<sup>10</sup>. The direct effects of warming on decomposers may be either mitigated or amplified by restructuring of the plant community, soil food web, and soil temperature regime. However, uncertainty regarding the strength of positive versus negative feedbacks that couple increased shrub dominance, decomposers and their shared soil environment complicates the ability to project future tundra carbon dynamics<sup>11</sup>.

Warming promotes woody-shrub growth in Arctic systems<sup>11,12</sup>. Shrubs trap snow, creating warmer winter soil temperatures that stimulate decomposers and support shrub dominance further<sup>3</sup>. Shrubs also increase summer shade, thereby reducing decomposer activity by

cooling soils<sup>13</sup>. Whether the winter-warming or summer-cooling effects will dominate temperature-driven changes in decomposition, or whether they will effectively neutralize each other is unknown. Shrub expansion also alters both the nature of carbon inputs and where in the soil profile they occur. Relatively labile shrub leaf litter<sup>14,15</sup> replaces slowly decomposing mosses at the soil surface, whereas greater root biomass may increase labile carbon inputs deeper in the soil<sup>14,15</sup>. Such changes in the nature and location of carbon inputs will alter the activity and community structure of decomposers further<sup>9</sup>. Labile, nitrogen-rich litter favours a rapidly cycling community dominated by bacteria that promotes nitrogen availability<sup>16</sup>, whereas lignified woody litter selects for a more slowly cycling fungus-dominated food web<sup>9</sup>.

As mineral interactions can protect soil carbon physically, control over decomposition in mineral soils is different to that in overlying organic soils<sup>17</sup>. Greater inputs of labile plant-derived carbon into the mineral horizon may stimulate decomposer communities<sup>18,19</sup>, while increasing aggregate formation that can stabilize carbon and limit loss<sup>20</sup>. In parallel with these biotic feedbacks, increased thaw depth under warmer conditions will increase SOM availability in the mineral horizon, stimulating decomposers<sup>2</sup>. However, summer shade may ultimately reduce thaw depth, despite greater annual average soil temperatures<sup>13,21</sup>. The largest fraction of the permafrost soil carbon is contained in the mineral soil<sup>1</sup>, and therefore these feedbacks can substantially alter tundra carbon storage over time.

We investigated the effects of two decades of summer warming on a moist acidic tussock (MAT) tundra ecosystem, to evaluate how warming-induced feedbacks affect Arctic carbon cycling. We sampled the longest-running whole-system tundra warming experiment: an *in situ* greenhouse experiment that was started in 1989 at the US Arctic Long Term Ecological Research site (68° 38'N, 149° 34'W). Twenty years of warming increased plant carbon storage and altered both plant community structure and the belowground food web; however, there was no net change in total soil carbon ( $P = 0.5$ ) or nitrogen stocks

**Table 1 | Effects of greenhouse warming on tundra carbon and nitrogen pools after 20 years of treatment**

Soil characteristic	Surface organic soil		Deep organic soil		Mineral soil	
	Control	Greenhouse	Control	Greenhouse	Control	Greenhouse
Carbon ( $\text{g m}^{-2}$ )	1,026.9 $\pm$ 101.2	929.1 $\pm$ 101.8	4,547.2 $\pm$ 1064.4	3,906.6 $\pm$ 638.0	<b>6,381.7 <math>\pm</math> 1411.1</b>	<b>8,342.3 <math>\pm</math> 786.5</b>
Nitrogen ( $\text{g m}^{-2}$ )	26.5 $\pm$ 4.1	27.5 $\pm$ 3.1	163.5 $\pm$ 47.2	145.5 $\pm$ 18.1	318.0 $\pm$ 75.5	376.9 $\pm$ 29.3
Carbon:nitrogen ratio	<b>42.2 <math>\pm</math> 1.7</b>	<b>35.8 <math>\pm</math> 2.6</b>	31.5 $\pm$ 1.4	28.4 $\pm$ 1.3	<b>20.4 <math>\pm</math> 0.4</b>	<b>22.4 <math>\pm</math> 0.6</b>
% Carbon	44.6 $\pm$ 1.5	38.2 $\pm$ 2.9	30.4 $\pm$ 2.5	28.3 $\pm$ 2.0	7.2 $\pm$ 1.5	8.7 $\pm$ 1.8
% Nitrogen	1.1 $\pm$ 0.04	1.1 $\pm$ 0.08	0.97 $\pm$ 0.067	1.1 $\pm$ 0.11	0.37 $\pm$ 0.083	0.39 $\pm$ 0.076
Bulk density ( $\text{g soil per cm}^3$ )	0.05 $\pm$ 0.006	0.06 $\pm$ 0.02	0.14 $\pm$ 0.04	0.2 $\pm$ 0.04	1.0 $\pm$ 0.11	0.8 $\pm$ 0.1
Sampling depth (cm)	5 $\pm$ 0	5 $\pm$ 0	10.1 $\pm$ 1.5	11.2 $\pm$ 1.2	17.0 $\pm$ 1.6	15.7 $\pm$ 0.8
Percentage soil moisture ( $\text{g H}_2\text{O per g dry soil}$ )	443.5 $\pm$ 40.5	531.6 $\pm$ 65.2	391.8 $\pm$ 27.4	446.1 $\pm$ 82.3	75.5 $\pm$ 14.9	107.9 $\pm$ 24.1

Numbers represent the mean  $\pm$  1 s.e. Soil pools include surface litter and fine roots. Surface organic soil, less than 5 cm in depth from the surface; deep organic soil, the soil between the surface organic and the mineral horizon. The mineral horizon was sampled to permafrost. Means that significantly differ from each other (within horizons,  $n = 4$ ;  $\alpha \leq 0.1$ ) are indicated in bold. Mineral-horizon soil carbon and nitrogen stocks were also analysed using a modified equivalent soil-mass method. The maximum sampled control mineral-horizon soil mass among the four replicates was defined as the initial soil mass; all greenhouse and control mineral-horizon soil-sample replicates were set as equivalent to that mass and soil carbon stocks were recalculated. The equivalent soil-mass greenhouse mineral-horizon carbon stock was approximately 28% greater than control (control average = 7451.0  $\pm$  1043, greenhouse average = 9514.8  $\pm$  996,  $P = 0.06$ ).

<sup>1</sup>Department of Ecology, Evolution and Marine Biology, University of California Santa Barbara, Santa Barbara, California 93108, USA. <sup>2</sup>Natural Resource Ecology Laboratory, Colorado State University, Fort Collins, Colorado 80523, USA. <sup>3</sup>Department of Ecosystem Science and Sustainability, Colorado State University, Fort Collins, Colorado 80523, USA. <sup>4</sup>Department of Biology, University of Texas at Arlington, Arlington, Texas 76019, USA. <sup>5</sup>The Ecosystems Center, Marine Biological Laboratory, Woods Hole, Massachusetts 02543, USA.



**Table 2 | Effects of warming on plant community**

(a), Effects of greenhouse warming on per cent cover of tundra plants and litter

Plant	Year	Treatment	Control (1998, 2008)	Greenhouse (1998, 2008)
<i>Betula nana</i>	*	*	13.5 ± 2.0, 16.0 ± 2.3	16.6 ± 3.3, 31.0 ± 6.8
<i>Salix pulchra</i>	NS	NS	3.7 ± 1.2, 3.8 ± 1.5	5.7 ± 1.4, 5.3 ± 1.6
<i>Eriophorum vaginatum</i>	NS	*	10 ± 1.9, 19 ± 2.5	11.2 ± 2.8, 15 ± 4.2
<i>Carex</i> spp.	NS	***	3.0 ± 0.74, 3.0 ± 0.72	0.5 ± 0.25, 1.0 ± 0.38
<i>Vaccinium vitis-idea</i>	**	NS	14.0 ± 1.6, 11.3 ± 0.28	18.8 ± 0.4, 9.1 ± 2.3
<i>Rhododendron tomentosum</i>	NS	NS	8.3 ± 1.5, 10.6 ± 1.6	13.8 ± 2.0, 9.8 ± 0.86
<i>Cassiope tetragona</i>	*	***	7.0 ± 0.97, 5.7 ± 1.8	3.4 ± 1.2, 0.69 ± 0.23
<i>Empetrum nigrum</i>	**	NS	4.5 ± 0.26, 3.4 ± 0.6	4.6 ± 0.4, 2.3 ± 0.58
<i>Rubus chamaemorus</i>	NS	*	8.9 ± 2.9, 6.8 ± 1.6	13.1 ± 1.9, 11 ± 1.9
<i>Pedicularis lapponica</i>	NS	*	0.63 ± 0.26, 0.19 ± 0.05	0.12 ± 0.09, 0.013 ± 0.013
Lichen	NS	****	4.6 ± 1.2, 3 ± 0.27	0.075 ± 0.052, 0.036 ± 0.36
Moss	****	***	18.7 ± 2.9, 7.3 ± 0.98	8.2 ± 1.1, 2.7 ± 0.56
Litter	****	*	0.24 ± 0.14, 6.3 ± 0.8	2.6 ± 1.0, 8.8 ± 0.75
Deciduous shrub	*	*	17.2 ± 3.0, 20.4 ± 3.0	22.4 ± 3.3, 36.7 ± 7.0
Evergreen shrub	*	NS	34.1 ± 2.3, 31.2 ± 4.0	41.2 ± 3.7, 22.1 ± 3.7
Graminoid	*	*	13.4 ± 1.1, 21.9 ± 2.3	11.7 ± 2.9, 16.4 ± 4.5
Forb	NS	NS	10.9 ± 3.0, 7.8 ± 1.6	13.3 ± 1.9, 11.2 ± 1.9

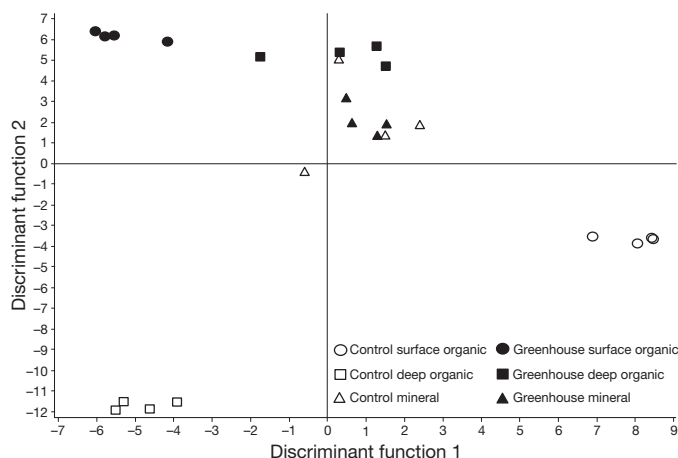
(b), Effects of 14 years of greenhouse warming (2002 harvest) on tundra plant biomass

Plant biomass (g dry weight per m <sup>2</sup> )	Treatment	Control	Greenhouse
Total aboveground (including mosses and lichens)	**	474.9 ± 32.6	748.7 ± 75.9
Vascular aboveground	***	369.5 ± 26.0	720.7 ± 85.9
Vascular belowground	*	438.3 ± 88.7	712.4 ± 70.6
Total aboveground and vascular belowground	**	913.1 ± 118.9	1,524.0 ± 112.6
Deciduous shrub	*	218.7 ± 51.8	551.2 ± 119.1
Evergreen shrub	*	324.5 ± 20.5	532.2 ± 80.8
Graminoid	NS	227.0 ± 57.5	179.0 ± 92.9
Forb	***	37.6 ± 12.3	170.7 ± 44.4
Litter and standing dead	*	569.3 ± 134	758.4 ± 171.4
Moss	**	75.5 ± 10.8	16.3 ± 4.4
Lichen	*	29.9 ± 6.4	11.8 ± 6.7

Numbers represent the mean ± 1 s.e. of plant percentage cover (sampled in 1998 and 2008), and plant biomass (sampled in 2002). \* $P < 0.1$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$  and \*\*\*\* $P < 0.0001$  for means that differ from each other significantly at  $\alpha \leq 0.1$  ( $n = 4$ ).

( $P = 0.7$ ; Table 1 and Supplementary Fig. 1). Although warming significantly increased surface litter inputs (Table 2a) and probably also increased their quality<sup>16</sup>, the greatest biogeochemical changes were at depth, in the mineral soil. The greenhouse treatment increased mineral soil carbon by 31% ( $P = 0.06$ ) and its carbon:nitrogen ratio by 9.8% ( $P = 0.05$ ), whereas the surface organic soil carbon:nitrogen ratio decreased by 15% ( $P = 0.02$ ; Table 1).

The stability of the tundra soil's overall carbon and nitrogen stocks under sustained warming contrasted with the vegetation (Table 2) and soil trophic structure (Fig. 1, Table 3 and Supplementary Tables 1–3). The greenhouse treatment affected the plant community structure



**Figure 1 | A canonical discriminant analysis plot reveals the loss of trophic heterogeneity across soil horizons in the greenhouse treatment relative to control conditions.** CDA biota groupings included total protozoa (amoebae, flagellates and ciliates were grouped) and did not include bacteria and fungi (which constitute a disproportionate majority of food-web biomass) ( $\mu\text{g}$  carbon per g soil; Wilks' lambda  $< 0.0001$ ;  $n = 4$ ; Table 3).

across all major MAT growth forms. The percentage cover of the deciduous shrub *Betula nana* increased by 94%, litter cover increased by 40%, and lichens and mosses declined by 99% and 63%, respectively. After 14 years of warming, greenhouse plant carbon stocks had increased by 50% (control =  $446.5 \pm 56.8 \text{ g C m}^{-2}$ , greenhouse =  $668.7 \pm 39.9 \text{ g C m}^{-2}$ ;  $P = 0.02$ ), whereas greenhouse plant nitrogen did not significantly respond (control =  $10.1 \pm 1.3 \text{ g N m}^{-2}$ , greenhouse =  $13.5 \pm 0.84 \text{ g N m}^{-2}$ ;  $P = 0.12$ ). The greenhouse effect on plant biomass was consistent with the percentage cover after 10 and 20 years of warming. The soil food web was also altered by the greenhouse treatment. The trophic structure of control MAT soil is dominated by fungi and is stratified by depth in biomass, diversity, connectance and linkage density, with each horizon possessing a distinct structure. The greenhouse treatment reduced the heterogeneity of the soil food web among horizons. In the control soil, food webs within each horizon differed from each other. Within the greenhouse treatment, only the soil communities within the surface organic (0 to 5 cm in depth) and mineral horizons differed (Fig. 1 and Table 3).

The surface organic soil carbon and nitrogen stocks were unaffected by warming although the carbon:nitrogen ratio declined (Table 1),

**Table 3 | Mahalanobis distances by horizon and treatment**

Horizon or treatment	Mahalanobis distances by horizon (within treatment)			Mahalanobis distances by treatment (within horizon)
	Surface organic versus deep organic	Surface organic versus mineral	Deep organic versus mineral	
Control	243.4**	131**	234.1**	-
Greenhouse	74.27	115.4**	44.1	-
Surface organic	-	-	-	275.6***
Deep organic	-	-	-	322.2***
Mineral	-	-	-	0.1

CDA units are squared. \*\* $P < 0.05$ , \*\*\* $P < 0.01$ .

**Table 4 | Effects of greenhouse warming on tundra microbial biomass, activity and extractable nutrient pools during the growing season**

Soil characteristic	Surface organic soil		Deep organic soil		Mineral soil	
	Control	Greenhouse	Control	Greenhouse	Control	Greenhouse
Microbial biomass carbon ( $\mu\text{g C per g soil}$ )	6,050 $\pm$ 1237	6,315 $\pm$ 1366	3,137 $\pm$ 496	3,810 $\pm$ 667	151 $\pm$ 44	176 $\pm$ 31
Microbial biomass nitrogen ( $\mu\text{g N per g soil}$ )	605 $\pm$ 106	563 $\pm$ 127	285 $\pm$ 63	307 $\pm$ 90	<b>5.78 <math>\pm</math> 1.55</b>	<b>13.0 <math>\pm</math> 3.86</b>
Fungal: bacterial biomass ratio	181 $\pm$ 53	220 $\pm$ 53	<b>858 <math>\pm</math> 305</b>	<b>163 <math>\pm</math> 69</b>	115 $\pm$ 37	76 $\pm$ 70
SIR active microbial biomass ( $\mu\text{g C-CO}_2 \text{ per g soil per day}$ )	<b>6,011 <math>\pm</math> 733</b>	<b>4,584 <math>\pm</math> 664</b>	1,604 $\pm$ 434	1,848 $\pm$ 351	<b>59.9 <math>\pm</math> 7.26</b>	<b>83.0 <math>\pm</math> 4.28</b>
Carbon mineralization (14-day incubation; $\mu\text{g C-CO}_2 \text{ per g soil per day}$ )	31.83 $\pm$ 1.6	27.15 $\pm$ 11.6	12.37 $\pm$ 2.2	15.46 $\pm$ 3.4	<b>0.43 <math>\pm</math> 0.05</b>	<b>0.64 <math>\pm</math> 0.04</b>
Extractable organic nitrogen ( $\text{mg m}^{-2}$ )*	3.93 $\pm$ 0.6	5.43 $\pm$ 1.3	25.16 $\pm$ 4.6	27.95 $\pm$ 8.8	84.31 $\pm$ 10.2	71.71 $\pm$ 9.8
Extractable organic carbon ( $\text{mg m}^{-2}$ )*	111.07 $\pm$ 18.3	138.11 $\pm$ 37.1	610.94 $\pm$ 192.3	917.60 $\pm$ 57.2	<b>1,343.14 <math>\pm</math> 160.2</b>	<b>1,092.11 <math>\pm</math> 97.1</b>

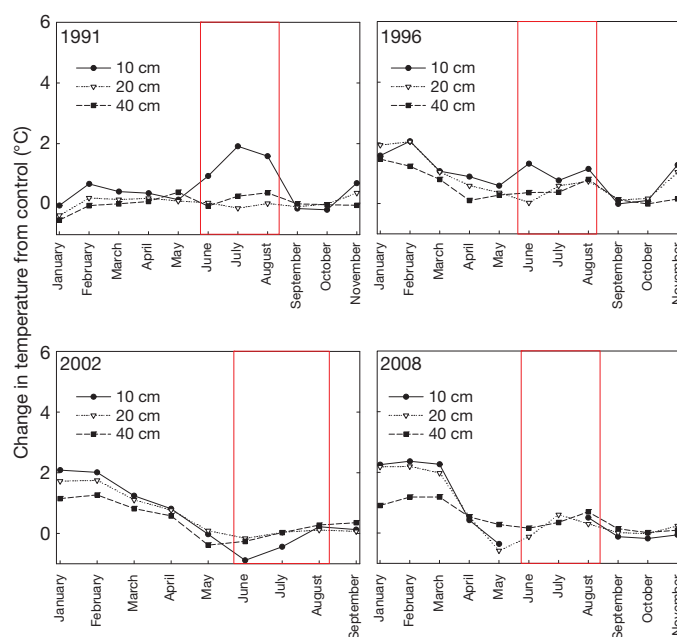
\* Scaling to  $\text{m}^{-2}$  uses average plot bulk density and volume. Numbers represent the mean  $\pm$  1 s.e. of soil microbial biomass, activity and extractable nutrient pools. Means that differ from each other significantly at  $\alpha \leq 0.1$  (within horizons;  $n = 4$ ) are indicated in bold. C-CO<sub>2</sub>, carbon derived from or originating from CO<sub>2</sub>.

reflecting a decline in carbon-rich mosses and lichens (Table 2) and increased inputs of leaf litter from *B. nana* and *Rubus chamaemorus* with lower carbon:nitrogen ratios<sup>22</sup>. We observed significant reductions in the greenhouse surface organic soil's root-feeding ( $P = 0.07$ ) and predator nematode biomass ( $P = 0.07$ ), as well as active microbial biomass (measured by substrate induced respiration (SIR); Table 4). This re-alignment of the trophic pathways and overall suppression of the surface organic decomposer food web activity highlights that changes in plant litter inputs alone do not account for soil-community responses to warming. It also challenges our hypothesis that soil nitrogen enrichment coupled with a decline in relatively recalcitrant matter correlates with more rapid carbon-cycling in the organic soil horizons. In the deep organic horizon (5 cm below the surface to mineral soil horizon), both bacterivores ( $P = 0.09$ ) and the ratio of fungal to bacterial biomass ( $P = 0.08$ ) declined, indicating that warming restricted both the root- and detritus-fed trophic pathways.

The altered soil-community structure and activity in the organic horizons may be driven in part by shifted greenhouse soil-temperature patterns, which are consistent with the snow-trap warming model of tundra shrub expansion (Fig. 2 and Supplementary Table 4). The greenhouse treatment initially warmed the soil so that its temperature was up to 2 °C higher than ambient temperature during the summer<sup>23</sup>. Although the summertime greenhouse soil warming declined over the course of the experiment<sup>23</sup>, by 1996 a winter warming effect had developed in the mineral soil that consistently exceeded the direct effects on summer soil temperature (Supplementary Table 4). We propose that at the onset of the experiment, soil respiration increased, similar to an increase in soil respiration that was observed 3.5 years into another warming experiment nearby<sup>5</sup> and in meta-analyses of short-term tundra responses to warming<sup>12</sup>. Declining summertime organic soil respiration (the dominant period of tundra soil activity<sup>24</sup>), coupled with greater litter inputs, may have offset initial warming-driven carbon losses, causing no net change in organic horizon carbon stocks.

Unlike the overlying organic horizons, the greenhouse treatment stimulated mineral soil-decomposer activity; microbial biomass nitrogen more than doubled, SIR biomass increased by 39%, and incubation-based carbon mineralization was approximately 1.5 times greater than ambient levels, whereas extractable organic carbon was reduced by 19% (Table 4). This may reflect increased labile carbon consumption by the more active decomposer community. The greenhouse food web became more homogenized across depth relative to control conditions, owing to a vertical redistribution of functional groups, including the presence of phytophagous nematodes and oribatid mites (not found in the mineral soil under control conditions), and the loss of non-oribatid mites, in the greenhouse mineral soil (Fig. 1, Table 3, and Supplementary Tables 1 and 2). Thus, although the detritus-fed pathway underwent a transition, the root-derived trophic pathway expanded.

The stimulation of the greenhouse mineral soil food web is probably caused by increases in late August-thaw depth (Supplementary Fig. 2), in annual soil temperature and in the number of plant-derived resources at depth. Root-derived carbon is a primary resource for soil food webs<sup>25</sup> but roots (which in MAT extend into the mineral soil<sup>15</sup>) also contribute substantially to stable SOM pools<sup>26</sup>. Increases in the volume of litter that moves deeper owing to cryoturbation<sup>3</sup>, and greater fine-root biomass at the interface of the organic and mineral horizons may have promoted the changes observed in the greenhouse mineral soil. Although we did not measure fine-root biomass in 2008, belowground vascular biomass had increased by 2002 (Table 2b), and *B. nana* root production is stimulated by warm conditions<sup>14</sup>. Furthermore, *B. nana* dominance increases overall rooting in the deep organic horizon<sup>14</sup>. We



**Figure 2 | A discontinuous record of the difference between greenhouse treatment and control soil.** Temperature probes up to approximately 10, 20 and 40 cm in depth, from 1991 to 2008. Monthly averages of daily temperatures were used in each block. Vertical lines indicate when the greenhouse is actively warming. Only in 1991 and 1996 are summer greenhouse soil temperatures significantly higher than control temperatures ( $P < 0.1$ ;  $n = 1$ ). Winter (January to March) temperatures do not initially differ between greenhouse and control; however, from 1996 to 2008, winter greenhouse soil temperatures exceeded ambient conditions in all horizons and exceeded the observed summer temperature differences ( $P < 0.0001$ ; Supplementary Table 4).

propose that warming stimulated greater rhizome and root biomass in the greenhouses' deep organic and mineral soils, which increased the carbon-rich root exudates and leachates moving into the mineral horizon in these highly water-saturated soils.

Long-term tundra warming increased plant biomass carbon and the dominance of woody plants, altered the structure and activity of the soil food web, and changed both the vertical distribution and turnover of soil carbon without affecting overall soil carbon storage. We infer from this that 20 years of warming has promoted net ecosystem carbon storage. Our results also show that deeper mineral soils are susceptible to coupled biotic–abiotic effects driven by warming over decades. Although increased decomposer activity did not offset increased carbon inputs in the mineral soil, incubation studies suggest that labile carbon limits tundra mineral-soil-decomposer activity<sup>19</sup>. Thus, although greater carbon availability at depth may initially increase carbon storage, it remains uncertain whether the ecosystem response observed after 20 years of warming reflects a continued trajectory of increased net carbon storage or a transient state in which an activated decomposer system will ultimately outpace carbon inputs. As such, identifying the mechanisms under which warming stimulates and regulates tundra decomposer activity at depth—where the majority of permafrost soil carbon is stored—remains a pressing challenge.

## METHODS SUMMARY

**Site description.** The study was established in 1989 in MAT near Toolik Lake, Alaska, the United States (68° 38'N, 149° 34'W), and is described in detail elsewhere<sup>15,23</sup>. MAT vegetation is similar to approximately  $0.9 \times 10^6 \text{ km}^2$  of tussock tundra found globally<sup>27</sup>.

**Experimental design.** The study consists of four blocks containing control and greenhouse plots. The greenhouses are made of wooden frames ( $2.5 \times 5 \times 1.5 \text{ m}$ ) and 0.15-mm polyethylene sheets, and are erected yearly, immediately after snow-melt. In July 2008, we sampled four randomly selected subplots per plot ( $n = 4$ )<sup>28</sup>. Green litter was removed, and organic monoliths ( $8 \times 8 \text{ cm}$ ) were collected and separated into surface organic (0–5 cm), deep organic (5 cm to mineral horizon) and mineral soil (sampled to the permafrost). Plant, soil and fauna traits were measured using standard methods<sup>28–30</sup>.

**Statistical analyses.** Data were tested for normality. A blocked analysis of variance (ANOVA), repeated measures ANOVA followed by a Holm–Bonferroni correction for multiple comparisons across years, or a Wilcoxon signed-rank test was used to compare means with a two-tailed  $t$ -test at  $\alpha \leq 0.1$ ; the fixed effects were treatment, year and their interaction (as appropriate), separated by horizon. Plant cover data were arcsine-square-root transformed and plant biomass data were natural-log-transformed. Food web data were log-transformed and analysed using ANOVA (comparing horizons across treatment and testing for a depth effect within treatment, and then using Tukey's honestly significant difference (HSD) for multiple comparisons) and canonical discriminant analysis (CDA). Food-web biomass estimates were the CDA interval variables; treatment and soil horizon combinations constituted the classification variables. The model was run with bacteria and fungi, with similar effects. A soil food web connectedness description based on functional groups was used to compare horizons within and across treatments.

**Full Methods** and any associated references are available in the online version of the paper.

**Received 27 November 2012; accepted 28 March 2013.**

**Published online 15 May 2013.**

1. Tarnocai, carbon *et al.* Soil organic carbon pools in the northern circumpolar permafrost region. *Glob. Biogeochem. Cycles* **23**, GB2023 (2009).
2. Schuur, E. A. G. *et al.* The effect of permafrost thaw on old carbon release and net carbon exchange from tundra. *Nature* **459**, 556–559 (2009).
3. Sturm, M. *et al.* Winter biological processes could help convert arctic tundra to shrubland. *Bioscience* **55**, 17–26 (2005).
4. Shaver, G. R. *et al.* Carbon turnover in Alaskan tundra soils: effects of organic matter quality, temperature, moisture and fertilizer. *J. Ecol.* **94**, 740–753 (2006).
5. Hobbie, S. E. & Chapin, F. S. The response of tundra plant biomass, aboveground production, nitrogen, and  $\text{CO}_2$  flux to experimental warming. *Ecology* **79**, 1526–1544 (1998).
6. Oberbauer, S. F. *et al.* Tundra  $\text{CO}_2$  fluxes in response to experimental warming across latitudinal and moisture gradients. *Ecol. Monogr.* **77**, 221–238 (2007).

7. Natali, S. M., Schuur, E. A. G. & Rubin, R. L. Increased plant productivity in Alaskan tundra as a result of experimental warming of soil and permafrost. *J. Ecol.* **100**, 488–498 (2012).
8. Hodkinson, I. D. *et al.* Global change and Arctic ecosystems: conclusions and predictions from experiments with terrestrial invertebrates on Spitsbergen. *Arct. Alp. Res.* **30**, 306–313 (1998).
9. Deslippe, J. R., Hartmann, M., Simard, S. W. & Mohn, W. W. Long-term warming alters the composition of Arctic soil microbial communities. *FEMS Microbiol. Ecol.* **82**, 303–315 (2012).
10. McKane, R. B. *et al.* Climatic effects on tundra carbon storage inferred from experimental data and a model. *Ecology* **78**, 1170–1187 (1997).
11. Myers-Smith, I. H. *et al.* Shrub expansion in tundra ecosystems: dynamics, impacts and research priorities. *Environ. Res. Lett.* **6**, 045509 (2011).
12. Rustad, L. E. *et al.* A meta-analysis of the response of soil respiration, net nitrogen mineralization, and aboveground plant growth to experimental ecosystem warming. *Oecologia* **126**, 543–562 (2001).
13. Blok, D. *et al.* The response of Arctic vegetation to the summer climate: relation between shrub cover, NDVI, surface albedo and temperature. *Environ. Res. Lett.* **6**, 035502 (2011).
14. Sullivan, P. F. *et al.* Climate and species affect fine root production with long-term fertilization in acidic tussock tundra near Toolik Lake, Alaska. *Oecologia* **153**, 643–652 (2007).
15. Clemmensen, K. E., Michelsen, A., Jonasson, S. & Shaver, G. R. Increased ectomycorrhizal fungal abundance after long-term fertilization and warming of two arctic tundra ecosystems. *New Phytol.* **171**, 391–404 (2006).
16. Wallenstein, M. D., McMahon, S. & Schimel, J. Bacterial and fungal community structure in Arctic tundra tussock and shrub soils. *FEMS Microbiol. Ecol.* **59**, 428–435 (2007).
17. Oades, J. M. Soil organic matter and structural stability: mechanisms and implications for management. *Plant Soil* **76**, 319–337 (1984).
18. Loya, W. M., Johnson, L. C. & Nadelhoffer, K. J. Seasonal dynamics of leaf- and root-derived C in arctic tundra mesocosms. *Soil Biol. Biochem.* **36**, 655–666 (2004).
19. Lavoie, M., Mack, M. C. & Schuur, E. A. G. Effects of elevated nitrogen and temperature on carbon and nitrogen dynamics in Alaskan arctic and boreal soils. *J. Geophys. Res.* **116**, G03013 (2011).
20. Six, J., Bossuyt, H., Degryze, S. & Denef, K. A history of research on the link between (micro)aggregates, soil biota, and soil organic matter dynamics. *Soil Tillage Res.* **79**, 7–31 (2004).
21. Anisimov, O., a., Shiklomanov, N. I. & Nelson, F. E. Global warming and active-layer thickness: results from transient general circulation models. *Glob. Planet. Change* **15**, 61–77 (1997).
22. Hobbie, S. E., Gough, L. & Shaver, G. R. Species compositional differences on different-aged glacial landscapes drive contrasting responses of tundra to nutrient addition. *J. Ecol.* **93**, 770–782 (2005).
23. Deslippe, J. R., Hartmann, M., Mohn, W. W. & Simard, S. W. Long-term experimental manipulation of climate alters the ectomycorrhizal community of *Betula nana* in Arctic tundra. *Glob. Change Biol.* **17**, 1625–1636 (2011).
24. Fahnestock, J. T., Jones, M. H. & Welker, J. M. Wintertime  $\text{CO}_2$  efflux from Arctic soils: implications for annual carbon budgets. *Glob. Biogeochem. Cycles* **13**, 775–779 (1999).
25. Pollierer, M. M., Langel, R., Körner, C., Maraun, M. & Scheu, S. The underestimated importance of belowground carbon input for forest soil animal food webs. *Ecol. Lett.* **10**, 729–736 (2007).
26. Loya, W. M. Pulse-labeling studies of carbon cycling in arctic tundra ecosystems: contribution of photosynthates to soil organic matter. *Glob. Biogeochem. Cycles* **16**, 1101 (2002).
27. Oechel, W. C. *et al.* Recent change of Arctic tundra ecosystems from a net carbon dioxide sink to a source. *Nature* **361**, 520–523 (1993).
28. Mack, M. C., Schuur, E. A. G., Bret-Harte, M. S., Shaver, G. R. & Chapin, F. S., III. Ecosystem carbon storage in arctic tundra reduced by long-term nutrient fertilization. *Nature* **431**, 440–443 (2004).
29. Fierer, N. & Schimel, J. P. Effects of drying–rewetting frequency on soil carbon and nitrogen transformations. *Soil Biol. Biochem.* **34**, 777–787 (2002).
30. Bloem, J. Fluorescent staining of microbes for total direct counts. *Mol. Microbial Ecol. Manual* 1–12 (1995).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** This research was supported by a DOE Global Change Education Program Graduate Fellowship, a Leal Anne Kerry Mertes scholarship, and Explorer's Club grant to S.A.S., NSF OPP-1023524 to J.P.S., NSF OPP-0425606 and NSF OPP-0909441 to J.C.M., NSF OPP-0425827 and NSF OPP-0909507 to L.G., and the Arctic LTER program NSF-DEB 1026843 to G.R.S. We thank J. Laundre for temperature and thaw depth data. We also thank three anonymous reviewers, C. D'Antonio, J. King, and S. Viswanathan for comments that greatly improved this manuscript.

**Author Contributions** S.A.S., J.C.M., L.G., J.P.S. and G.R.S. conceived the study and designed scientific objectives. J.C.M. and R.T.S. collected the soil food web data, L.G. and G.R.S. collected the plant data, S.A.S. and J.P.S. collected the soil chemistry data. S.A.S., J.C.M., L.G. and R.T.S. carried out statistical analyses. S.A.S. wrote the paper. J.P.S., J.C.M., L.G. and G.R.S. provided textual edits and all authors commented on the analysis and presentation of the data.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.A.S. (sistla@lifesci.ucsb.edu).



## METHODS

**Experimental treatment.** The greenhouse treatment reduces photosynthetically active radiation and direct input from precipitation, but does not negatively influence plant growth or affect soil moisture (as far as it was possible to detect this), and uneven microtopography enables air circulation beneath the greenhouse bases<sup>15,23</sup>. The greenhouse air temperature is elevated by 2 °C on average, and soil temperature increased by more than 1 °C on average (up to 40 cm in depth)<sup>23</sup>. Soil temperature is measured at a depth of approximately 10, 20 and 40 cm using a Campbell CR21x data logger that recorded two profiles of soil temperature per treatment in one block with copper or constantin thermocouple wires. Soil-temperature data are patchy because of field-equipment failures and irreconcilable movement of the probes with freeze–thaw cycles. The probability of making a type I error ( $\alpha \leq 0.1$ ) was chosen to balance the potential for making type II versus type I errors, given the study's small replicate number ( $n = 4$ ).

**Soil harvest and analyses of soil carbon and nitrogen pools.** Live coarse roots were removed from soil subsamples by hand, and the volume and mass of rocks were determined and subtracted from the volume and mass of soil in the laboratory. The remaining soil was subsampled for gravimetric water content, and carbon and nitrogen concentrations. Soil subsamples were dried to a constant mass and carbon and nitrogen concentrations were determined using a Fisons NA1500 carbon and nitrogen analyser. The mass of carbon and nitrogen in each horizon (g carbon per m<sup>2</sup> and g nitrogen per m<sup>2</sup>) was calculated by multiplying the percentage of carbon or nitrogen per subsample with the corresponding quadrat subsample bulk density and volume, and dividing this value by the subsample area. Treatment effects on the percentage of carbon or nitrogen in soil, and on the mass of carbon and nitrogen in horizons were analysed at the block level using the average of the four block subsamples. Subsamples from the four individual quadrats, separated by horizon, were composited (aggregated) at the block level by hand for further biogeochemical analyses at the block level<sup>28</sup>.

**Vegetative cover and biomass.** Vegetative cover was quantified visually in eight permanent 1-m<sup>2</sup> quadrats per plot. A plant harvest in July 2002 used four subplots per block, with vegetation (excluding roots) cut through the mineral-soil surface (including rhizomes that extended into the mineral horizon)<sup>28</sup>.

**Carbon mineralization and SIR microbial biomass.** Compositing soils (approximately 8 g wet weight for the surface organic soil and approximately 10 g wet weight for the deep organic and mineral horizons) were weighed out into airtight incubation vessels and incubated at approximately 5 °C (similar to *in situ* soil conditions) for 14 days to capture carbon-mineralization rates immediately after soil harvest. Cumulative carbon mineralization (CO<sub>2</sub> accumulation) was measured by drawing headspace air samples from the incubation vessels using airtight glass syringes and injecting each sample into an infrared gas analyser (IRGA; LI-COR 6252). The carbon-mineralization rate was calculated as the change in headspace CO<sub>2</sub> concentrations (in µg C-CO<sub>2</sub> (carbon originating or derived from CO<sub>2</sub>)) per g soil (dry weight equivalent) per day over the 14-day incubation period. Microbial biomass at the end of the carbon-mineralization incubation period was estimated by SIR<sup>31,32</sup>. Samples were amended with autolysed yeast extract (a 10-ml solution of 3 g yeast extract dissolved into 250 ml milli-Q H<sub>2</sub>O) in airtight tubes and placed horizontally on a shaker (100 r.p.m.) for 6 h. Headspace CO<sub>2</sub> concentrations were monitored after 3 and 6 h. The average respiration rate (µg C-CO<sub>2</sub> per g soil per h) over the incubation period is an index of the SIR-responsive microbial biomass.

**Extractable microbial-biomass carbon and nitrogen.** Chloroform-extractable microbial biomass was determined as described elsewhere<sup>29</sup> on approximately 8 g wet weight of surface organic and approximately 10 g wet weight of deep organic and mineral composited soil, immediately after soil collection. Microbial-biomass carbon or nitrogen (extractable plus non-extractable biomass) was calculated by dividing the measured extractable biomass values by correction factors of 0.45 for

carbon<sup>33</sup> and 0.54 for nitrogen<sup>34</sup>. Extractable inorganic nitrogen was analysed using the 0.5 M K<sub>2</sub>SO<sub>4</sub> extracts on a Lachat autoanalyzer (NH<sub>4</sub><sup>+</sup> using Lachat method no. 31-107-06-5-A; NO<sub>3</sub><sup>-</sup> using Lachat method no. 12-107-04-1-B).

**Direct counts of bacteria and fungi.** Bacterial and fungal biomass were estimated using a direct count method<sup>30,35</sup>. All finished samples were stored at 4 °C until direct counts could be made. Bacterial direct counts were made using a confocal microscope at ×1,500 magnification. Biomass conversions were made assuming an average dry weight of  $6.65 \times 10^{-13}$  g per bacterial cell<sup>36</sup>. Fungal hyphal length was estimated using a confocal microscope at ×400 magnification by counting the number of times hyphae crossed an ocular lens grid. Hyphal length was estimated using the equation  $R = \frac{\pi NA}{2H}$ , where  $R$  is the total hyphal length,  $N$  is the number of times hyphae crossed the horizontal lines on the grid,  $A$  is the area of one slide well, and  $H$  is the total length of the horizontal grid lines. Biomass conversions were made assuming  $2.3 \times 10^{-6}$  g m<sup>-1</sup> of hyphae.

**Protozoan biomass.** Protozoan densities were estimated using a most-probable-number (MPN) technique<sup>37</sup>, with a starting serial dilution of 10 g of sieved soil in 90 ml of sterile de-ionized water, with tenfold dilutions to 10<sup>-6</sup> ml. Four 0.5-ml subsamples were taken from each dilution in the series and individually transferred to a well in a standard 24-well plate. A 50-µl suspension of the bacterium *Escherichia coli* was transferred to each well. Plates were incubated at 26 °C for 3 days, after which individual wells were observed using an inverted compound microscope at ×100 magnification. Wells were scored for the presence or absence of 'ciliates', 'flagellates' and 'amoebae'. The densities of total protozoa, and the individual subgroupings were estimated from the frequency distribution of positive reads across the dilution series using the MPN estimation program provided by the US Environmental Protection Agency.

**Invertebrate biomass.** Soil fauna were isolated from fresh soils and enumerated as described previously<sup>38</sup>. Soil nematodes were isolated using Baermann funnels. Microfauna were heat extracted into a solution of 90% ethanol and 1% glycerine from soils using Tullgren funnels. The intact samples were wrapped individually in cheesecloth, weighed, placed over a funnel and heated with a 9-W incandescent light for 5 days. The dried soils were re-weighed, and the arthropods were sorted into taxonomic groupings and functional groups. Enchytraeids, ciliates, rotifers, tardigrades, nematodes and microarthropods (collembola, mites, copepods and insects) were counted from 5-g subsamples of sieved soil immersed in de-ionized water using a dissecting microscope.

31. West, A. W. & Sparling, G. P. Modifications to the substrate-induced respiration method to permit measurement of microbial biomass in soils of differing water contents. *J. Microbiol. Methods* **5**, 177–189 (1986).
32. Anderson, J. P. E. & Domsch, K. H. A physiological method for the quantitative measurement of microbial biomass in soils. *Soil Biol. Biochem.* **10**, 215–221 (1978).
33. Beck, T. *et al.* An inter-laboratory comparison of ten different ways of measuring soil microbial biomass C. *Soil Biol. Biochem.* **29**, 1023–1032 (1997).
34. Brookes, P. C., Landman, A., Prudenmaltese, G. & Jenkinson, D. S. Chloroform fumigation and the release of soil nitrogen: a rapid direct extraction method to measure microbial biomass nitrogen in soil. *Soil Biol. Biochem.* **17**, 837–842 (1985).
35. Frey, S., Elliott, E. & Paustian, K. Bacterial and fungal abundance and biomass in conventional and no-tillage agroecosystems along two climatic gradients. *Soil Biol. Biochem.* **31**, 573–585 (1999).
36. Ilic, B. *et al.* Single cell detection with micromechanical oscillators. *J. Vac. Sci. Technol. B* **19**, 2825–2828 (2001).
37. Darbyshire, J. F., Wheatley, R. E., Greaves, M. P. & Inkson, R. H. E. A rapid micromethod for estimating bacterial and protozoan populations in soil. *Rev. Ecol. Biol. Sol* **11**, 465–475 (1974).
38. Gough, L., Moore, J. C., Shaver, G. R., Simpson, R. T. & Johnson, D. R. Above- and belowground responses of arctic tundra ecosystems to altered soil nutrients and mammalian herbivory. *Ecology* **93**, 1683–1694 (2012).

# Synthetic analog computation in living cells

Ramiz Daniel<sup>1,2,3</sup>, Jacob R. Rubens<sup>2,3,4</sup>, Rahul Sarpeshkar<sup>1,3,4,5,6,7\*</sup> & Timothy K. Lu<sup>2,3,4,5,6,8\*</sup>

**A central goal of synthetic biology is to achieve multi-signal integration and processing in living cells for diagnostic, therapeutic and biotechnology applications<sup>1</sup>. Digital logic has been used to build small-scale circuits, but other frameworks may be needed for efficient computation in the resource-limited environments of cells<sup>2,3</sup>. Here we demonstrate that synthetic analog gene circuits can be engineered to execute sophisticated computational functions in living cells using just three transcription factors. Such synthetic analog gene circuits exploit feedback to implement logarithmically linear sensing, addition, ratiometric and power-law computations. The circuits exhibit Weber's law behaviour as in natural biological systems<sup>4</sup>, operate over a wide dynamic range of up to four orders of magnitude and can be designed to have tunable transfer functions. Our circuits can be composed to implement higher-order functions that are well described by both intricate biochemical models and simple mathematical functions. By exploiting analog building-block functions that are already naturally present in cells<sup>3,5</sup>, this approach efficiently implements arithmetic operations and complex functions in the logarithmic domain. Such circuits may lead to new applications for synthetic biology and biotechnology that require complex computations with limited parts, need wide-dynamic-range biosensing or would benefit from the fine control of gene expression.**

In natural biological systems, digital behaviour is appropriate in settings where decision making is necessary, such as in developmental circuits<sup>6</sup>. The digital approach is an abstraction of graded analog functions, where values above a threshold are classified as '1' and values below this threshold are classified as '0' (Fig. 1a). Digital computation using synthetic gene circuits has included switches, counters, logic gates, classifiers and edge detectors (see references 28–40 in Supplementary Information). However, given that there is often unwanted crosstalk amongst synthetic devices<sup>7</sup> and cellular resource limitations<sup>2,3</sup>, it may be challenging to scale digital logic functions to the level needed for complex computations in living cells. Analog functions can be found in natural biological systems, where they enable graded responses to environmental signals<sup>1,8</sup>. For example, neurons can implement both digital and analog computation<sup>9</sup>. Furthermore, electronic circuits that perform analog computation on logarithmically transformed signals have been used in commercially valuable electronic chips for several decades. The thermodynamic Boltzmann exponential equations that describe electron flow in electronic transistors are strikingly mathematically similar to the thermodynamic Boltzmann exponential equations that describe molecular flux in chemical reactions<sup>5</sup>. Their similarities suggest that log-domain analog computation in electronics may be mapped to log-domain analog computation in chemistry, and vice versa<sup>5</sup>. Because analog computation exploits powerful biochemical mathematical basis functions<sup>3,10</sup> that are naturally present over the entire continuous range of input operation (Fig. 1a), they are an advantageous alternative to digital logic when resources of device count, space, time or energy are constrained<sup>2,3</sup>.

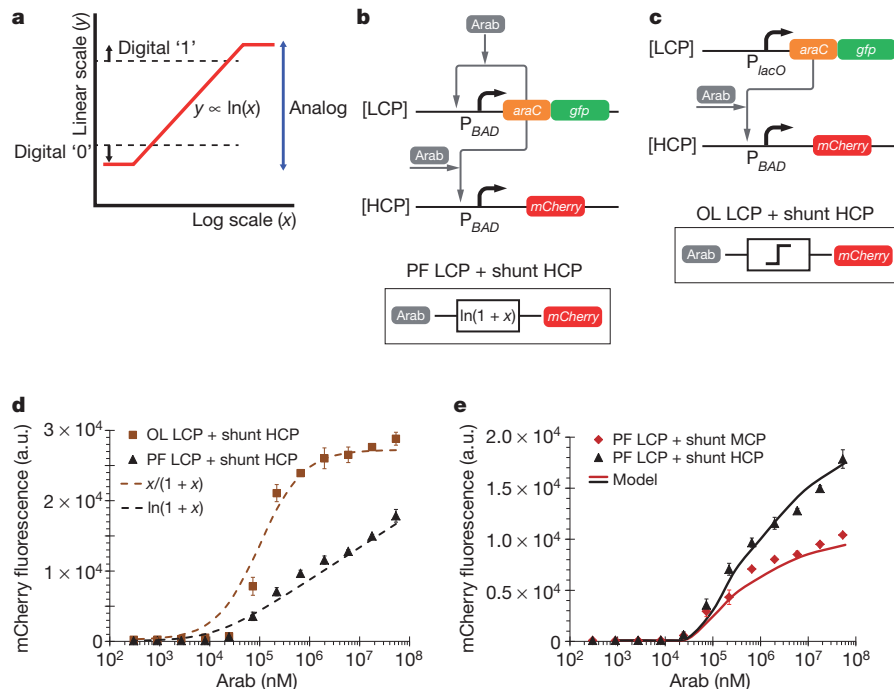
We first created an analog synthetic gene circuit motif that performs positive, wide-dynamic-range logarithmic transduction of inducer concentration inputs to fluorescent protein outputs (Fig. 1b). The resulting transfer functions have a broad region of linearity on a semi-log plot (log-linear). Logarithmic functions can enable intensity-independent responses and can compress a large input dynamic range into a smaller, manageable output dynamic range. A logarithmic function naturally implements Weber's law behaviour, which states that the ratio between the perceptual change in a signal divided by its background level is a constant, resulting in the detection of fold changes rather than absolute levels<sup>4</sup>. Weber's law is approximately true within molecular signalling networks and in describing human perception of sound intensity, light intensity and weight<sup>4</sup>.

Our wide-dynamic-range, positive-logarithm circuit consists of a positive-feedback component located on a low-copy-number plasmid (LCP) and a 'shunt' component located on a high-copy-number plasmid (HCP). The LCP comprises promoters which generate transcription factors that bind to the promoters, thus increasing the generation of transcription factors by means of a positive-feedback loop and alleviating saturation of the interaction between inducers and transcription factors. The HCP comprises promoters that bind and shunt away the generated transcription factors, effectively redirecting some of them away from the LCP and reducing saturation of the interaction between transcription factors and DNA. The HCP shunt also affects the effective strength of the LCP positive-feedback loop. The HCP generates a reporter or protein output that is different from the transcription factor output (Fig. 1b). We initially applied this 'positive-feedback/shunt' (PFS) circuit motif to the AraC transcriptional activator, which is induced by arabinose and activates the  $P_{BAD}$  promoter (Fig. 1b and Supplementary Fig. 6).

The AraC-based PFS circuit had an input–output transfer function that was well fitted by a simple mathematical function of the form  $\ln(1+x)$ , which is a first-order approximation of the Hill function for small values of  $x$ , where  $x$  is a scaled version of the input concentration (Fig. 1d and Supplementary Information, section 4). Furthermore, this circuit had a wide input dynamic range of more than three orders of magnitude, where the dynamic range is taken to be the span of inputs over which the output is well fitted by  $\ln(x)$  (Fig. 1d and Supplementary Fig. 18). The simple logarithmic mathematical functions that describe our wide-dynamic-range circuits are useful for designing higher-order functions. The wide-dynamic-range behaviour of our circuit was especially striking when compared with the narrow dynamic range of the open-loop control, which has a shunt motif but no positive-feedback motif. This 'open-loop/shunt' motif is shown in Fig. 1c, d and in Supplementary Fig. 16. When the shunt plasmid in the PFS motif contains a  $P_{lux}$  promoter rather than a  $P_{BAD}$  promoter, wide-dynamic-range logarithmic operation for the AraC-based circuit is also absent (Supplementary Fig. 17). These controls indicate the importance of graded positive feedback, implemented here with the PFS motif, in achieving wide-dynamic-range operation.

<sup>1</sup>Analog Circuits and Biological Systems Group, Research Lab of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. <sup>2</sup>Synthetic Biology Group, Research Lab of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. <sup>3</sup>Synthetic Biology Center, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. <sup>4</sup>MIT Microbiology Program, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. <sup>5</sup>Department of Electrical Engineering & Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. <sup>6</sup>MIT Computational and Systems Biology Program, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. <sup>7</sup>MIT Biophysics Program, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. <sup>8</sup>Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

\*These authors contributed equally to this work.



**Figure 1 | Positive-feedback linearization of gene circuits for wide-dynamic-range analog computation.** **a**, Synthetic analog gene circuits use the inherent continuous behaviour of biochemical reactions to perform computations and implement mathematical functions over a wide dynamic range, whereas digital circuits abstract this behaviour into discrete '0's and '1's. **b**, The AraC-based, wide-dynamic-range, positive-logarithm circuit logarithmically transforms input inducer concentrations into output protein levels. This topology involves a transcriptional positive-feedback (PF) loop on an LCP, which alleviates saturated binding of inducer to transcription factor, along with a shunt HCP containing transcription-factor-binding sites, which alleviates saturation of DNA-binding sites. Arab, arabinose. **c**, The open-loop (OL) control consists of AraC–GFP expression from a  $P_{lacO}$  promoter with no positive feedback. **d**, Arabinose-to-mCherry transfer function. The positive-

feedback LCP with a shunt HCP (black triangles) implements a wide-dynamic-range, positive-slope logarithm circuit with an input dynamic range of more than three orders of magnitude. It is well fitted by a mathematical function of the form  $\ln(1+x)$ , where  $x$  is a scaled version of the input inducer concentration. In contrast, the open-loop LCP with a shunt HCP (orange squares) has a narrow dynamic range and is well fitted by a Hill function. a.u., arbitrary units. **e**, Comparing the positive-feedback LCP with a shunt MCP (red diamonds) and the positive-feedback LCP with a shunt HCP (black triangles; data from **d**) demonstrates the importance of the shunt plasmid in making possible the wide-dynamic-range operation. Solid lines indicate modelling results of the detailed biochemical model (Supplementary Information). The errors (s.e.m.) shown in the figures are derived from three flow cytometry experiments, each of which involved  $n = 50,000$  events.

To gain deeper insights into the mechanisms that may give rise to logarithmically linear transfer functions, we built detailed biochemical models that capture the effects of inducer/transcription-factor binding, transcription-factor/DNA binding, the PFS circuit topology and protein degradation (Fig. 1e and Supplementary Information, section 2). Using a consistent set of model parameters that differ only on the basis of the various circuit topologies (for example in plasmid copy number), our biochemical models accurately capture the behaviour of our multiple circuits (Figs 1 and 2 and Supplementary Fig. 31). A minimal biochemical model, which incorporates only the basic effects of graded positive feedback, also shows linearization (Supplementary Fig. 30). Indeed, our scheme for widening the log-linear dynamic range of operation by means of graded positive feedback is conceptually general and applies to both genetic and electronic circuits: expansive hyperbolic-sine-based linearization of compressive hyperbolic-tangent-based functions in log-domain electronic circuits<sup>11</sup> is analogous to the use of expansive positive-feedback linearization of compressive biochemical binding functions in log-domain genetic circuits.

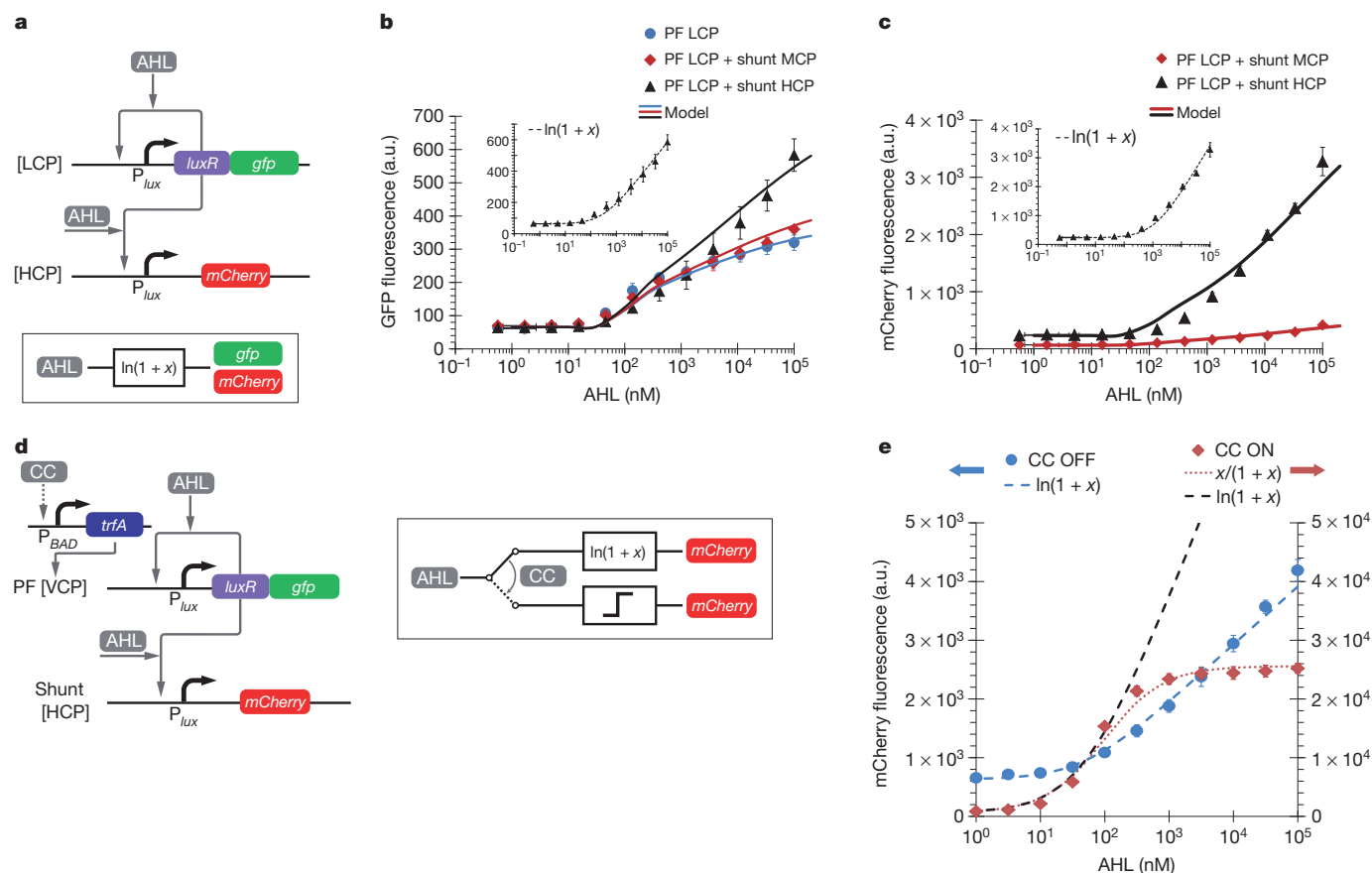
We generalized these findings to the quorum-sensing transcriptional activator LuxR, which is induced by acyl homoserine lactone (AHL) and activates the promoter  $P_{lux}$  (Fig. 2a). As fluorescent outputs of this circuit, we fused green fluorescent protein (GFP) to the carboxy terminus of LuxR and used an HCP  $P_{lux}$ –mCherry shunt. The LuxR PFS circuit also had an input dynamic range of more than three orders of magnitude (Fig. 2b) and performed robustly over multiple time points (Supplementary Fig. 26). This input dynamic range was significantly greater than that achieved with LuxR–GFP positive feedback alone or with LuxR–GFP positive feedback with a medium-copy-number plasmid (MCP)

shunt (Fig. 2b). The output of the shunt plasmid (mCherry) had similar properties and thus can also be used for computation (Fig. 2c). As in the AraC-based circuits (Fig. 1), both detailed biochemical models (Fig. 2b, c), where the only varying parameter was the plasmid copy number, and the simple  $\ln(1+x)$  mathematical function (Fig. 2b, c, insets) captured the behaviour of the LuxR-based circuits.

The behaviour of the PFS circuit motif can be dynamically tuned by changing the relative copy numbers of the positive-feedback and shunt plasmids. We demonstrated such tuning by combining an HCP shunt with a variable-copy-number plasmid (VCP), based on a pBAC/oriV vector<sup>12</sup>, carrying the positive-feedback component (Fig. 2d). When the VCP was induced to a high-copy-number state, the circuit had a narrow dynamic range of about two orders of magnitude and was poorly fitted by a  $\ln(1+x)$  function but could be fitted by a 'digital-like' Hill function (Fig. 2e). When the VCP was in a low-copy-number state, the circuit behaved in an analog fashion, followed a  $\ln(1+x)$  mathematical relationship and had a broad dynamic range of nearly four orders of magnitude. Such tuning demonstrates the importance of the relative copy numbers of the positive-feedback and shunt plasmids in enabling wide-dynamic-range logarithmic operation. It also provides a mechanism for actively changing circuit behaviour between analog and digital modes and shows that the PFS circuit motif can be reliably used in different *Escherichia coli* strain backgrounds.

Our analog computation modules can be composed into more complex circuits for higher-order functions, as shown in Supplementary Information, section 10, and Supplementary Fig. 31. There we show how to compose a wide-dynamic-range PFS logarithm motif with a LacI repressor module to create a wide-dynamic-range negative





**Figure 2 | Analog behaviour versus digital.** **a**, The LuxR-based, wide-dynamic-range, positive-logarithm circuit. **b**, The AHL-to-GFP transfer function for a positive-feedback LCP (blue circles), a positive-feedback LCP with a shunt MCP (red diamonds) and a positive-feedback LCP with a shunt HCP (black triangles). The positive-feedback LCP with a shunt HCP implements a wide-dynamic-range, positive-slope logarithm circuit. Solid lines indicate modelling results of the detailed biochemical model; the inset shows the fit of a mathematical function of the form  $\ln(1+x)$ . **c**, The AHL-to-mCherry transfer function for a positive-feedback LCP with a shunt MCP (red diamonds) and a positive-feedback LCP with a shunt HCP (black triangles). The positive-feedback LCP with a shunt HCP implements a wide-dynamic-range, positive-slope logarithm circuit. Solid lines indicate modelling results of the detailed biochemical model; the inset shows the fit of a mathematical function of the form  $\ln(1+x)$ . **d**, Placing the positive-feedback loop on a VCP

logarithm motif of the form  $-\ln(1+x)$ . The latter motif functions over four orders of magnitude and is well predicted by detailed biochemical models and by compositions of empirical fits.

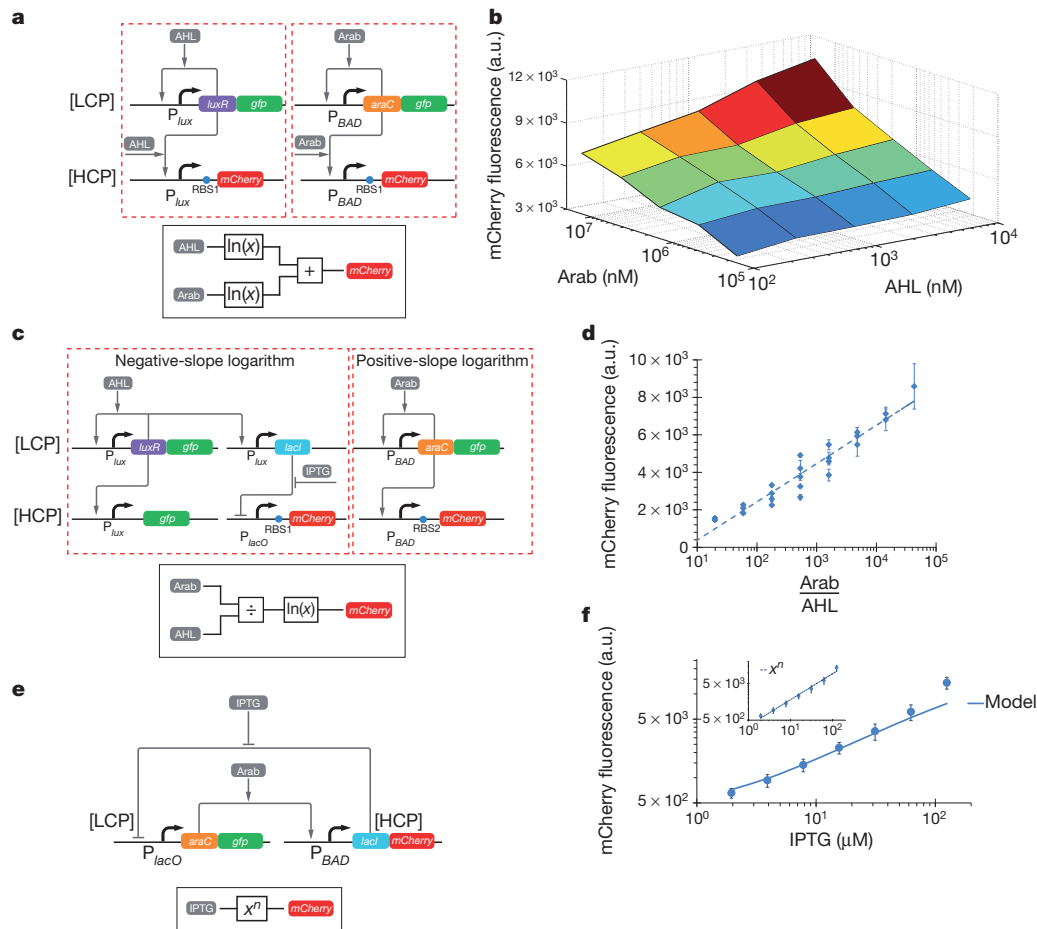
As an application of synthetic analog genetic circuits, we created an adder circuit that sums its protein inputs, which are logarithmically transformed representations of inducer molecules (Fig. 3a). Such summation is achieved by combining two parallel, wide-dynamic-range, positive-logarithm circuits that accept different input molecules (for example AHL and arabinose) but produce a common output molecule (Fig. 3a). The adder exhibited log-linear behaviour over a range of two orders of magnitude (Fig. 3b and Supplementary Fig. 21). Because log-linear addition of two inputs effectively results in the logarithm of their product, an analog product is equivalent to a 'soft AND' operation, that is, to a graded multiplication rather than to a one-bit digital multiplication. Thus, the data of Fig. 3b have similarities to that obtained from digital AND circuits, except that the overall function is more graded in nature.

We also developed a PFS circuit that calculates the log-transformed ratio of two different input inducers (Fig. 3c). Ratiometric calculations are useful in biological systems, because they enable the normalization of measurements, comparisons between variables and decisions based

allows the dynamic adjustment of AHL-to-mCherry transfer functions between analog and digital behaviours using a CopyControl (CC) induction solution. The VCP is normally maintained at low copy numbers and can be induced to higher copy numbers through CopyControl-mediated expression of replication protein TrfA from a promoter integrated into the genome of EPI300 cells<sup>12</sup>. **e**, At high copy numbers (CC ON, red diamonds), the circuit behaves in a digital-like fashion. The dotted red line is a Hill function fit to the digital-like curve. The dashed black line reveals that the digital-like curve is not well fitted by a  $\ln(1+x)$  function. At low copy numbers (CC OFF, blue circles), the circuit behaves in an analog fashion with a wide dynamic range. The dashed blue line indicates that the PFS positive logarithm is well fitted by a  $\ln(1+x)$  function. The errors (s.e.m.) in the figures are derived from three flow cytometry experiments, each involving  $n = 50,000$  events.

on competing inputs. Our ratiometer circuit was built by combining a wide-dynamic-range, negative-logarithm circuit and a wide-dynamic-range, positive-logarithm circuit that accept different input molecules but produce a common output molecule (Fig. 3d). This circuit essentially calculates the difference between the log-transformed outputs of the two inputs (subtraction in the log domain). By tuning the ribosome-binding sequences of the negative- and positive-logarithm circuits such that the magnitudes of their slopes are similar, the resulting mathematical function is the log-transformed ratio of the two inputs and functions over four orders of magnitude of this ratio. Our wide-dynamic-range ratiometer enables the concept of pH, which measures the logarithmic concentration ratio of  $H^+$  with respect to an absolute value, to be generalized to the concept of 'pRATIO', which may be useful for measuring the logarithmic concentration ratio of one input with respect to another input.

In addition to logarithmically linear addition and ratiometric calculations, synthetic analog circuits in living cells can implement power laws. To implement such laws, we introduced an inducer/transcription-factor binding function into a strong negative-feedback loop that includes two stages of amplification (Fig. 3e). In the first stage, araC-GFP produced from PlacO on a LCP activates lacI-mCherry produced



**Figure 3 | Log-domain analog computation.** Complex analog computation can be implemented by composing synthetic gene circuits. **a**, An adder is built by engineering two wide-dynamic-range, positive-slope logarithm circuits (modules outlined in red) to produce a common output, which is summed to yield the overall output. **b**, The adder circuit sums the logarithms of two inputs, AHL and arabinose, over about two orders of magnitude, to produce an output, mCherry. **c**, A division circuit, or ratiometer, is implemented when the slopes of the wide-dynamic-range, positive- and negative-logarithms (modules outlined in red) are closely matched by tuning their output ribosome-binding sites (RBSs). IPTG, isopropyl- $\beta$ -D-1-thiogalactopyranoside. **d**, The ratiometer performs a logarithmic transformation on the ratio of two inputs, arabinose

and AHL, over more than three orders of magnitude. IPTG was held constant at 1.5 mM. The dotted blue line indicates a log-linear fit. **e**, A negative-feedback loop with tunable feedback strength implements power-law functions. This circuit motif uses LacI-mCherry produced on an HCP to suppress the production of AraC-GFP on an LCP. When induced by arabinose, AraC-GFP enhances the production of LacI-mCherry. **f**, Power-law behaviour can be observed in the IPTG-to-mCherry transfer function, with arabinose constant at 4.6 mM. The solid line indicates modelling results of the detailed biochemical model; the inset shows the fit to a power law of the form  $x^{0.7}$ . The errors (s.e.m.) shown in the figures are derived from three flow cytometry experiments, each of which involved  $n = 50,000$  events.

from PBAD on a HCP. In the second stage, LacI-mCherry represses araC-GFP production at PlacO, completing the negative-feedback loop. The power-law nature of our circuit arises from the interactions of saturated-repressor polynomial functions and a linear activator polynomial function in a feedback loop (Supplementary Information, section 2). The power-law behaviour of our circuit extended over two orders of magnitude, was accurately predicted by our detailed biochemical models and was well matched by a simple power law,  $x^n$  (Fig. 3f). We have shown that powerful wide-dynamic-range analog computations can be performed with just a few biological parts in living cells. Four-input-bit, two-output-bit square roots can be calculated *in vitro* using 130 DNA strands within a DNA-based computation framework<sup>13</sup>. In comparison, our *in vivo*, analog power-law circuit exploits binding functions that are already present in the biochemistry and therefore requires only two transcription factors. Even one-bit full adders and subtractors in digital computation require several logic gates and, thus, numerous synthetic parts (refs 34, 35 and 37 Supplementary Information). Our wide-dynamic-range analog adders and ratiometers are inherently implemented by circuits that add flux to or subtract flux from a common output molecule, and can be constructed with no more than three transcription factors (Fig. 3).

In Supplementary Information, section 11, we outline potential applications of our efficient circuit motifs in biotechnology and medicine for fine control of gene expression or wide-dynamic-range sensing. Our wide-dynamic-range analog computation circuits can be further integrated with dynamical systems such as oscillators<sup>14–16</sup>, negative-feedback linearizing circuits<sup>17,18</sup>, endogenous circuits<sup>19</sup>, cell-cell communication<sup>20–22</sup> and cellular computers<sup>23</sup>, and can be implemented using RNA components<sup>24–26</sup>, synthetic transcriptional regulation<sup>27</sup> or protein-protein interactions<sup>28</sup>.

Using fundamental properties of the scaling laws of thermodynamic noise with temperature and molecular count, which hold in both biological and electronic systems, the advantages and disadvantages of analog computation versus digital have been analysed for neurobiological systems<sup>3</sup> and for systems in cell biology<sup>5</sup>. These results show that analog computation is more efficient than digital computation in part count, speed and energy consumption below a certain crossover computational precision. Although the exact crossover precision varies with the computation, both in electronics and in actual biological cells, the use of feedback loops, calibration loops, technological basis functions, redundancy, signal averaging and error-correcting topologies can push this crossover precision to higher values<sup>3</sup>. Alternatively,

for a given speed of operation, more energy must be expended in creating a higher molecular production rate, which leads to a higher molecular count and, thus, higher precision<sup>3,5</sup>. Trade-offs between error and resource use are therefore inherent in the design of synthetic circuits in living cells. To demonstrate the tunability of our analog circuits, we constructed an AraC PFS circuit with two  $P_{BAD}$  promoters on the shunt plasmid, leading to an increase in the log-linear gain of about twofold relative to the circuit's single- $P_{BAD}$  counterpart (Supplementary Fig. 25). We have also analysed the sensitivities of our circuits, defined as the fractional change in the output divided by the fractional change in the input, and found that they compare favourably to circuits operating with positive feedback only or in open-loop configurations (Supplementary Fig. 27).

We propose that an efficient and accurate computational approach to synthetic biological networks will ultimately integrate both analog and digital processing (a simple example of switched analog computation is shown in Supplementary Fig. 24). This mixed-signal approach can utilize analog or collective analog computation<sup>5</sup> (the latter with many interacting moderate-precision analog circuits) for front-end processing in concert with decision-making digital circuits; or it may use graded feedback for simultaneous analog and digital computation, as in neuronal networks in the brain<sup>29</sup>. Thus, future efforts can seek to integrate synthetic analog and digital computation in living cells to achieve enhanced computational power, efficiency, reliability and memory. Such mixed-signal processing would benefit from the development of circuits to convert signals from analog to digital, and vice versa<sup>5,30</sup>. These complementary approaches promise to broaden the computational power of synthetic biology and the range of applications that can be addressed using techniques from this emerging engineering field.

## METHODS SUMMARY

**Strains and plasmids.** All plasmids in this work were constructed using basic molecular cloning techniques (Supplementary Information). *Escherichia coli* 10β (araD139 Δ(ara-leu)7697 fhuA lacX74 galK (Φ80 Δ(lacZ)M15) mcrA galU recA1 endA1 nupG rpsL (StrR) Δ(mrr-hsdRMS-mcrBC)) or *E. coli* EPI300 (F- mcrA Δ(mrr-hsdRMS-mcrBC) Φ80dlacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara-leu)7697 galU galK λ- rpsL (StrR) nupG trfA tonA), where noted, were used as bacterial hosts for the circuits in Figs 1–3.

**Circuit characterization.** Overnight cultures of *E. coli* strains were grown from glycerol freezer stocks at 37 °C, in a VWR 1585 shaking incubator at 300 r.p.m., in 3 ml of Luria–Bertani–Miller medium (Fisher) with appropriate antibiotics: carbenicillin (50 μg ml<sup>-1</sup>), kanamycin (30 μg ml<sup>-1</sup>), chloramphenicol (25 μg ml<sup>-1</sup>). The inducers used were arabinose, isopropyl-β-D-1-thiogalactopyranoside and AHL 3OC6HSL (Sigma-Aldrich). Where appropriate, CopyControl<sup>12</sup> from Epicentre was added to overnight cultures at ×1 active concentration. Overnight cultures were diluted 1:100 into 3 ml fresh Luria–Bertani medium and antibiotics, and were incubated at 37 °C and 300 r.p.m. for 20 min. Cultures (200 μl) were then moved into 96-well plates, combined with inducers and incubated for 4 h 20 min in a VWR microplate shaker (37 °C, 700 r.p.m.), until they had an attenuation of  $D_{600\text{ nm}} \sim 0.6$ –0.8.

Cells were then diluted fourfold into a new 96-well plate containing fresh ×1 PBS and immediately assayed using a BD LSRFortessa high-throughput sampler. Detailed characterization methods can be found in Supplementary Information. At least 50,000 events were recorded in all experiments, and these data were then gated by forward scatter and side scatter using CYFLOGIC v.1.2.1 software (CyFlo). The geometric means of the gated fluorescence distributions were calculated using MATLAB. Fluorescence values were based on geometric means of flow cytometry populations from three experiments, each of which corresponded to 50,000 events, and the error bars represent standard errors of the mean.

**Full Methods** and any associated references are available in the online version of the paper.

Received 10 October 2012; accepted 3 April 2013.

Published online 15 May 2013.

- Chen, Y. Y., Galloway, K. E. & Smolke, C. D. Synthetic biology: advancing biological frontiers by building synthetic systems. *Genome Biol.* **13**, 240 (2012).
- Cardinale, S. & Arkin, A. P. Contextualizing context for synthetic biology: identifying causes of failure of synthetic biological systems. *Biotechnol. J.* **7**, 856–866 (2012).

- Sarpeshkar, R. Analog versus digital: extrapolating from electronics to neurobiology. *Neural Comput.* **10**, 1601–1638 (1998).
- Ferrell, J. E. Signaling motifs and Weber's law. *Mol. Cell* **36**, 724–727 (2009).
- Sarpeshkar, R. *Ultra Low Power Bioelectronics: Fundamentals, Biomedical Applications, and Bio-Inspired Systems* 651–694, 753–786 (Cambridge Univ. Press, 2010).
- Sprinzak, D. *et al.* Cis-interactions between Notch and Delta generate mutually exclusive signalling states. *Nature* **465**, 86–90 (2010).
- Canton, B., Labno, A. & Endy, D. Refinement and standardization of synthetic biological parts and devices. *Nature Biotechnol.* **26**, 787–793 (2008).
- Giorgetti, L. *et al.* Noncooperative interactions between transcription factors and clustered DNA binding sites enable graded transcriptional responses to environmental inputs. *Mol. Cell* **37**, 418–428 (2010).
- Clark, B. & Hausser, M. Neural coding: hybrid analog and digital signalling in axons. *Curr. Biol.* **16**, R585–R588 (2006).
- Daniel, R., Woo, S. S., Turicchia, L. & Sarpeshkar, R. in *Biomedical Circuits and Systems Conference (BioCAS 2011)* 333–336 (IEEE, 2011).
- Tavakoli, M. & Sarpeshkar, R. A sinh resistor and its application to tanh linearization. *IEEE J. Solid-State Circuits* **40**, 536–543 (2005).
- Wild, J., Hradecna, Z. & Szybalski, W. Conditionally amplifiable BACs: switching from single-copy to high-copy vectors and genomic clones. *Genome Res.* **12**, 1434–1444 (2002).
- Qian, L. & Winfree, E. Scaling up digital circuit computation with DNA strand displacement cascades. *Science* **332**, 1196–1201 (2011).
- Stricker, J. *et al.* A fast, robust and tunable synthetic gene oscillator. *Nature* **456**, 516–519 (2008).
- Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338 (2000).
- McMillen, D., Kopell, N., Hasty, J. & Collins, J. J. Synchronizing genetic relaxation oscillators by intercell signaling. *Proc. Natl Acad. Sci. USA* **99**, 679–684 (2002).
- Madar, D., Dekel, E., Bren, A. & Alon, U. Negative auto-regulation increases the input dynamic-range of the arabinose system of *Escherichia coli*. *BMC Syst. Biol.* **5**, 111 (2011).
- Nevozhay, D., Adams, R. M., Murphy, K. F., Josić, K. & Balázsi, G. Negative autoregulation linearizes the dose–response and suppresses the heterogeneity of gene expression. *Proc. Natl Acad. Sci. USA* **106**, 5123–5128 (2009).
- Shen-Orr, S. S., Milo, R., Mangan, S. & Alon, U. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature Genet.* **31**, 64–68 (2002).
- You, L., Cox, R. S., Weiss, R. & Arnold, F. H. Programmed population control by cell–cell communication and regulated killing. *Nature* **428**, 868–871 (2004).
- Tabor, J. J. *et al.* A synthetic genetic edge detection program. *Cell* **137**, 1272–1281 (2009).
- Tamsir, A., Tabor, J. J. & Voigt, C. A. Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'. *Nature* **469**, 212–215 (2011).
- Auslander, S., Auslander, D., Muller, M., Wieland, M. & Fussenegger, M. Programmable single-cell mammalian biocomputers. *Nature* **487**, 123–127 (2012).
- Isaacs, F. J. *et al.* Engineered riboregulators enable post-transcriptional control of gene expression. *Nature Biotechnol.* **22**, 841–847 (2004).
- Win, M. N. & Smolke, C. D. Higher-order cellular information processing with synthetic RNA devices. *Science* **322**, 456–460 (2008).
- Xie, Z., Wroblewska, L., Prochazka, L., Weiss, R. & Benenson, Y. Multi-input RNAi-based logic circuit for identification of specific cancer cells. *Science* **333**, 1307–1311 (2011).
- Khalil, A. *et al.* A synthetic biology framework for programming eukaryotic transcription functions. *Cell* **150**, 647–658 (2012).
- Dueber, J. E., Yeh, B. J., Chak, K. & Lim, W. A. Reprogramming control of an allosteric signaling switch through modular recombination. *Science* **301**, 1904–1908 (2003).
- Hahnloser, R. H. R., Sarpeshkar, R., Mahowald, M. A., Douglas, R. J. & Seung, H. S. Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. *Nature* **405**, 947–951 (2000).
- Lu, T. K., Khalil, A. S. & Collins, J. J. Next-generation synthetic gene networks. *Nature Biotechnol.* **27**, 1139–1150 (2009).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We would like to thank J. Nungesser for assistance with figures and members of the Lu and Sarpeshkar laboratories for discussions. This work was supported in part by a campus collaboration initiative from Lincoln Labs (R.D. and R.S.), the US National Science Foundation (R.D., J.R.R., R.S. and T.K.L.) under grant number 1124247, and the Office of Naval Research (J.R.R. and T.K.L.) under grant number N000141110725.

**Author Contributions** R.D., R.S. and T.K.L. designed the study. R.D. and J.R.R. performed experiments and collected data. R.D., J.R.R., R.S. and T.K.L. invented the analog circuit motifs. R.D., R.S. and T.K.L. developed the analog circuit motifs and associated models and simulations. All authors analysed the data, discussed results and wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.S. (rahuls@mit.edu) and T.K.L. (timlu@mit.edu).



## METHODS

**Strains and plasmids.** All plasmids in this work were constructed using basic molecular cloning techniques (Supplementary Information). *Escherichia coli* 10 $\beta$  (araD139  $\Delta$ (ara-leu)7697 fhuA lacX74 galK ( $\Phi$ 80  $\Delta$ (lacZ)M15) mcrA galU recA1 endA1 nupG rpsL (StrR)  $\Delta$ (mrr-hsdRMS-mcrBC)) or *E. coli* EPI300 (F- mcrA  $\Delta$ (mrr-hsdRMS-mcrBC)  $\Phi$ 80dlacZ $\Delta$ M15  $\Delta$ lacX74 recA1 endA1 araD139  $\Delta$ (ara, leu)7697 galU galK  $\lambda$ - rpsL (StrR) nupG trfA tonA), where noted, were used as bacterial hosts for the circuits in Figs 1–3.

**Circuit characterization.** Overnight cultures of *E. coli* strains were grown from glycerol freezer stocks in a shaking incubator at 37 °C and 300 r.p.m. in 3 ml of Luria–Bertani–Miller medium (Fisher), with appropriate antibiotics: carbenicillin (50  $\mu$ g ml<sup>−1</sup>), kanamycin (30  $\mu$ g ml<sup>−1</sup>), chloramphenicol (25  $\mu$ g ml<sup>−1</sup>). The inducers used were arabinose, IPTG and AHL 3OC6HSL (Sigma-Aldrich). Where appropriate, CopyControl<sup>12</sup> from Epicentre was added to overnight cultures at  $\times 1$  active concentration. Overnight cultures were diluted 1:100 into 3 ml fresh Luria–Bertani medium and antibiotics, and were incubated at 37 °C and 300 r.p.m. for 20 min. Cultures (200  $\mu$ l) were then moved into 96-well plates, combined with inducers and incubated for 4 h 20 min in a VWR incubating microplate shaker (37 °C, 700 r.p.m.) until they had an attenuation of  $D_{600\text{nm}} \sim 0.6$ –0.8.

Cells were then diluted fourfold into a new 96-well plate containing fresh  $\times 1$  PBS and immediately assayed using a BD LSRFortessa high-throughput sampler. Detailed characterization methods can be found in Supplementary Information. At least 50,000 events were recorded in all experiments, and these data were then gated by forward scatter and side scatter using CYFLOGIC v.1.2.1 software (CyFlo). The geometric means of the gated fluorescence distributions were calculated using MATLAB. Fluorescence values are based on geometric

means of flow cytometry populations from three experiments and the error bars represent standard errors of the mean.

When they varied, inducers were titrated across rows through serial dilutions. When an inducer was kept constant, it was added to cultures before they were moved to 96-well plates.

**Plate reader/FACS set-up.** For each experiment, fluorescence readings were taken on a BioTek Synergy H1 Microplate Reader using BioTek GEN5 software to determine the minimum and maximum expression levels for cultures in each 96-well plate. GFP fluorescence was quantified by excitation at a wavelength of 484 nm and emission at a wavelength of 510 nm. mCherry fluorescence was quantified by excitation at 587 nm and emission at 610 nm. The gain of the plate reader was automatically sensed and adjusted by the machine.

Cultures containing the minimum and maximum fluorescence levels, as determined by the plate reader, were used to calibrate the fluorescein isothiocyanate and PE-TexasRed filter voltages on a BD LSRFortessa high-throughput sampler to measure GFP and mCherry expression levels, respectively. The FACS voltages were adjusted using BD FACSDIVA software so that the maximum and minimum expression levels could be measured with the same voltage settings. Thus, consistent voltages were used across each entire experiment. The same voltages were used for subsequent repetitions of the same experiment. GFP was excited with a 488-nm laser and mCherry was excited with a 561-nm laser. Voltage compensation for fluorescein isothiocyanate and PE-TexasRed was not necessary for any experiment. Supplementary Information, section 12, provides our FACS data.

**Plasmid maps and construction.** Supplementary Information, sections 13 and 14, provide details regarding plasmid construction and plasmid maps, respectively. The references in Supplementary Information, section 15, provide details regarding the origin of the plasmids.

# BAF complexes facilitate decatenation of DNA by topoisomerase II $\alpha$

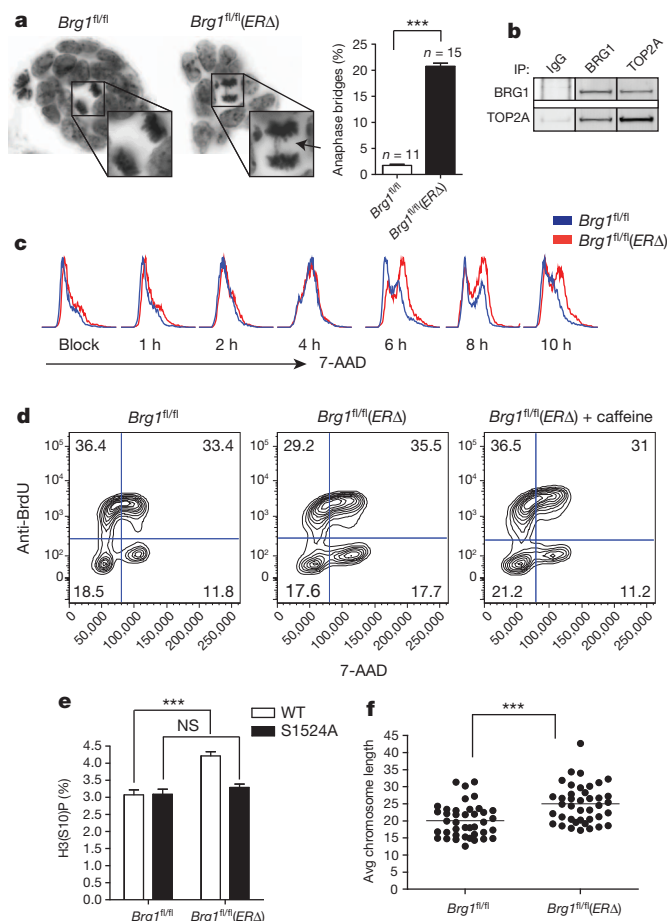
Emily C. Dykhuizen<sup>1,2\*</sup>, Diana C. Hargreaves<sup>1,2\*</sup>, Erik L. Miller<sup>1,3</sup>, Kairong Cui<sup>4</sup>, Andrey Korshunov<sup>5</sup>, Marcel Kool<sup>6</sup>, Stefan Pfister<sup>6,7</sup>, Yoon-Jae Cho<sup>8</sup>, Keji Zhao<sup>4</sup> & Gerald R. Crabtree<sup>1,2</sup>

Recent exon-sequencing studies of human tumours have revealed that subunits of BAF (mammalian SWI/SNF) complexes are mutated in more than 20% of all human malignancies<sup>1,2</sup>, but the mechanisms involved in tumour suppression are unclear. BAF chromatin-remodelling complexes are polymorphic assemblies that use energy provided by ATP hydrolysis to regulate transcription through the control of chromatin structure<sup>3</sup> and the placement of Polycomb repressive complex 2 (PRC2) across the genome<sup>4,5</sup>. Several proteins dedicated to this multisubunit complex, including BRG1 (also known as SMARCA4) and BAF250a (also known as ARID1A), are mutated at frequencies similar to those of recognized tumour suppressors. In particular, the core ATPase BRG1 is mutated in 5–10% of childhood medulloblastomas<sup>6–9</sup> and more than 15% of Burkitt's lymphomas<sup>10,11</sup>. Here we show a previously unknown function of BAF complexes in decatenating newly replicated sister chromatids, a requirement for proper chromosome segregation during mitosis. We find that deletion of *Brg1* in mouse cells, as well as the expression of BRG1 point mutants identified in human tumours, leads to anaphase bridge formation (in which sister chromatids are linked by catenated strands of DNA) and a G2/M-phase block characteristic of the decatenation checkpoint. Endogenous BAF complexes interact directly with endogenous topoisomerase II $\alpha$  (TOP2A) through BAF250a and are required for the binding of TOP2A to approximately 12,000 sites across the genome. Our results demonstrate that TOP2A chromatin binding is dependent on the ATPase activity of BRG1, which is compromised in oncogenic BRG1 mutants. These studies indicate that the ability of TOP2A to prevent DNA entanglement at mitosis requires BAF complexes and suggest that this activity contributes to the role of BAF subunits as tumour suppressors.

Using *Brg1*<sup>flxed/flxed</sup> (*Brg1*<sup>fl/fl</sup>) actin-*creER* mouse embryonic stem cells (ESCs), we observed that tamoxifen-induced deletion of *Brg1* (*Brg1*<sup>fl/fl</sup>(*ER* $\Delta$ )) results in the appearance of DNA bridges during anaphase (Fig. 1a). This phenotype is characteristic of a deficiency in TOP2A function, which normally resolves DNA catenanes that develop during transcription and replication<sup>12</sup>. We determined the frequency of anaphase bridges in *Brg1*<sup>fl/fl</sup>(*ER* $\Delta$ ) cells to be similar to that of cells deficient in other putative tumour suppressors that regulate TOP2A function, including BRCA1, RANBP2 and RECQL5 (refs 13–15) (Fig. 1a).

In previous studies, we recovered peptides from TOP2A in mass spectrometric analysis of endogenous BAF complexes<sup>16</sup>. Immunoprecipitation of BAF complexes with antibodies to BRG1 recovered TOP2A and, conversely, immunoprecipitation of TOP2A revealed BRG1 (Fig. 1b). This association is not dependent on DNA (Supplementary Fig. 1a, b). We detected this association in several additional cell types, including mouse embryonic fibroblasts (MEFs) and human embryonic kidney (HEK293T) cells (Supplementary Fig. 1c).

Failure of TOP2A to resolve catenated DNA leads to slow progression through the G2/M phase of the cell cycle<sup>17</sup>. To understand the mitotic defect in BRG1-deficient cells better, we synchronized *Brg1*<sup>fl/fl</sup> and



**Figure 1 | BRG1 associates with TOP2A and regulates its function. a**, Anaphase bridges in *Brg1*<sup>fl/fl</sup> and *Brg1*<sup>fl/fl</sup>(*ER* $\Delta$ ) ESCs. Data represent number of slides from four independent experiments  $\pm$  s.e.m. **b**, Co-immunoprecipitation of BRG1 and TOP2A from nuclear lysates. **c**, DNA content in *Brg1*<sup>fl/fl</sup> and *Brg1*<sup>fl/fl</sup>(*ER* $\Delta$ ) ESCs after release from double thymidine block. 7-AAD, 7-aminoactinomycin D. **d**, Cell cycle analysis of *Brg1*<sup>fl/fl</sup> and *Brg1*<sup>fl/fl</sup>(*ER* $\Delta$ ) ESCs with or without caffeine, an ATM/ATR inhibitor. **e**, Data represent five independent histone H3 serine 10 phosphorylation (H3(S10P)) cell cycle analyses  $\pm$  s.e.m. **f**, Data represent the mean of the average chromosome length per cell from 40 cells from metaphase spreads of *Brg1*<sup>fl/fl</sup> and *Brg1*<sup>fl/fl</sup>(*ER* $\Delta$ ) MEFs from two independent experiments.

<sup>1</sup>Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, California 94305, USA. <sup>2</sup>Department of Pathology, Stanford University School of Medicine, Stanford, California 94305, USA. <sup>3</sup>Department of Genetics, Stanford University School of Medicine, Stanford, California 94305, USA. <sup>4</sup>National Institutes of Health, Bethesda, Maryland 20892, USA. <sup>5</sup>CCU Neuropathology, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, Heidelberg 69120, Germany. <sup>6</sup>Division of Pediatric Neurooncology, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, Heidelberg 69120, Germany. <sup>7</sup>Department of Pediatric Oncology, University Hospital Heidelberg, Im Neuenheimer Feld 430, Heidelberg 69120, Germany. <sup>8</sup>Department of Neurology and Neurosurgery, Stanford University Medical School, Stanford, California 94305, USA.

\*These authors contributed equally to this work.

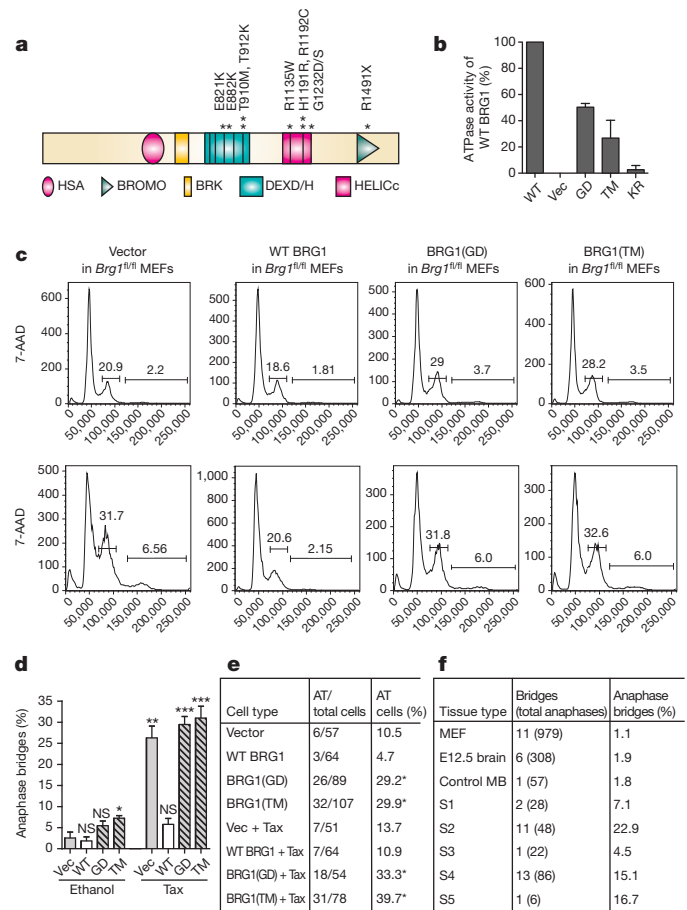
*Brg1*<sup>fl/fl</sup>(*ERA*) cells in G1/S using double thymidine block. After release, *Brg1*<sup>fl/fl</sup> and *Brg1*<sup>fl/fl</sup>(*ERA*) cells transited through the cell cycle at the same rate until reaching G2/M, where the *Brg1*<sup>fl/fl</sup>(*ERA*) cells exhibited a substantial delay (Fig. 1c). In asynchronously dividing cells, this delay resulted in a 1.5- to twofold increase in *Brg1*<sup>fl/fl</sup>(*ERA*) cells in G2/M (Fig. 1d), similar to the treatment of cells with the topoisomerase II inhibitor ICRF-193 (ref. 18). Caffeine, an inhibitor of ATM/ATR, forces cells through an ICRF-193-induced decatenation checkpoint<sup>18</sup> and similarly overrides the G2/M arrest in *Brg1*<sup>fl/fl</sup>(*ERA*) cells (Fig. 1d). Furthermore, expression of TOP2A(S1524A) (hereafter TOP2A(SA)), which fails to recruit MDC1 to chromatin upon initiation of the decatenation checkpoint<sup>19</sup>, alleviated the cell cycle delay, suggesting that *Brg1*<sup>fl/fl</sup>(*ERA*) cells arrest due to activation of the decatenation checkpoint (Fig. 1e and Supplementary Fig. 1d). Finally, chromosomes from *Brg1*<sup>fl/fl</sup>(*ERA*) cells are significantly longer than chromosomes from *Brg1*<sup>fl/fl</sup> cells (Fig. 1f and Supplementary Fig. 1e), a defect also observed in TOP2A-deficient cells owing to its role in chromosome condensation<sup>12,20</sup>. These data indicate that *Brg1* deletion resembles the mitotic defects observed with chemical inhibition and/or knockdown of TOP2A<sup>12,17,18,20</sup>.

We investigated oncogenic point mutants of BRG1 found in medulloblastoma and Burkitt's lymphoma (BRG1(G1232D), hereafter BRG1(GD); and BRG1(T910M), hereafter BRG1(TM))<sup>6–11</sup> by expressing Flag-tagged versions in *Brg1*<sup>fl/fl</sup> cells (Fig. 2a). The BRG1 mutants were incorporated into the BAF complex normally and did not alter the composition of the complex (Supplementary Fig. 2a). Although the T910M mutation is located in the ATP-binding pocket of BRG1, the G1232D mutation is downstream of the helicase superfamily carboxy-terminal (HELICc) domain and thus not obviously involved in ATP turnover (Fig. 2a). Subjecting BAF complexes containing BRG1(GD), BRG1(TM), wild-type BRG1 (hereafter WT BRG1) or BRG1(K798R) (hereafter BRG1(KR))<sup>21</sup>—the ATPase-dead point mutant of BRG1—to an assay for ATPase activity revealed that both cancer mutants are significantly compromised in ATPase activity, although not as profoundly as BRG1(KR) (Fig. 2b). BRG1(TM) is more severely compromised than BRG1(GD), which correlated with the viability of the respective cell line (Supplementary Fig. 2b).

Cells expressing BRG1(GD) and BRG1(TM), but not WT BRG1, display increases in the percentage of G2/M cells and the incidence of anaphase bridges similar to that of *Brg1*<sup>fl/fl</sup>(*ERA*) cells (Fig. 2c, d). Importantly, expression of the mutants in the presence of endogenous BRG1 gives similar increases, although less severe, in G2/M percentage and anaphase bridge incidence compared to vector cells (Fig. 2c, d). The dominant nature of these mutants on cell cycle and anaphase-bridge formation suggests that medulloblastomas with both heterozygous and homozygous mutations in BRG1 have these mitotic defects.

To explore whether the increase in anaphase bridges contributes to increased chromosome instability as it does in TOP2A-deficient cells<sup>12,17</sup>, we analysed ploidy in the BRG1 mutant cell lines. Expression of BRG1(GD) or BRG1(TM) results in a significant increase in cells with >4n DNA content in both ethanol- and tamoxifen-treated cells (Fig. 2c and Supplementary Fig. 2c). We also observed a significant increase in the number of BRG1(GD)- and BRG1(TM)-expressing cells with abnormal chromosome number in metaphase spreads from both ethanol- and tamoxifen-treated samples (Fig. 2e). These data indicate that the G1232D and T910M mutations in BRG1 can contribute to chromosome instability as a result of deficiencies in TOP2A function.

We collected several BRG1 mutant medulloblastomas to determine whether the effects of TOP2A deficiency can be observed in primary tumours. We observed an increased proportion of anaphase bridges in each of five histologic samples from BRG1 mutant tumours relative to controls, suggesting that these tumours have decatenation defects (Fig. 2f and Supplementary Fig. 2d). Aneuploidy is common in medulloblastoma and ranges from the partial gain or loss of single chromosomes to full tetraploidy<sup>8,22,23</sup>. However, a recent study showed that the relative rate of tetraploidy of 5 BRG1 mutant tumours was similar to that of 15



**Figure 2 | Expression of medulloblastoma-associated BRG1 mutants phenocopy TOP2A inhibition.** **a**, Somatic mutations in BRG1 found in medulloblastoma. BRK, Brm and KIS domain; BROMO, bromodomain; DEXD/H, DEAD/DEAH box helicase domain; HSA, helicase SANT-associated domain. **b**, The DNA-stimulated ATPase activity of BAF complexes from WT BRG1-, BRG1(GD)-, BRG1(KR)-, BRG1(TM)- and vector control-expressing *Brg1*<sup>fl/fl</sup>(*ERA*) cells  $\pm$  s.e.m. **c**, Cell cycle analysis of *Brg1*<sup>fl/fl</sup> MEFs expressing WT BRG1, BRG1(GD), BRG1(TM) or vector, treated with ethanol or tamoxifen (Tax). **d**, Cells were prepared as in **c** and the mean frequency of anaphase bridges  $\pm$  s.e.m. from three independent experiments was measured. **e**, Cells were prepared as in **c** and collected for metaphase spreads. The number of cells with greater than 40 chromosomes (AT) was quantitated from >50 cells. Significance was calculated relative to vector control ethanol-treated *Brg1*<sup>fl/fl</sup> cells where \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.0001$ . **f**, Various tissues were sectioned and scored for the number of anaphase bridges of total anaphases. MB, medulloblastoma; S, medulloblastoma subtype.

BRG1 wild-type tumours<sup>8</sup>. Additional sample characterization will be necessary to definitively assess whether BRG1 mutation causes mitotic defects through insufficient TOP2A function in medulloblastomas.

Microarray analysis of *Brg1*<sup>fl/fl</sup> and *Brg1*<sup>fl/fl</sup>(*ERA*) ESCs indicated that BRG1-dependent genes are not enriched for Gene Ontology terms related to DNA damage or repair<sup>24</sup>. Additionally, we found no alterations in the abundance, post-translational modifications, cellular localization, or *in vitro* enzymatic activity of TOP2A in *Brg1*<sup>fl/fl</sup>(*ERA*) cells (Supplementary Figs 3a–f and 4). To test whether purified BAF complexes could enhance the enzymatic activity of recombinant TOP2A, we used the standard *in vitro* kinetoplast DNA-based decatenation assay. Immobilized BAF complexes increased the enzymatic activity of TOP2A (Supplementary Fig. 5a); however, immunoprecipitated BRG1-mutant BAF complexes also enhanced TOP2A activity, as did immunoprecipitated SUZ12-containing PRC2 complexes, indicating a nonspecific activity on a bare DNA template<sup>25</sup> that does not reflect our *in vivo* observations. The BRG1 mutants did, however, reduce the association between TOP2A and chromatin, such that more TOP2A

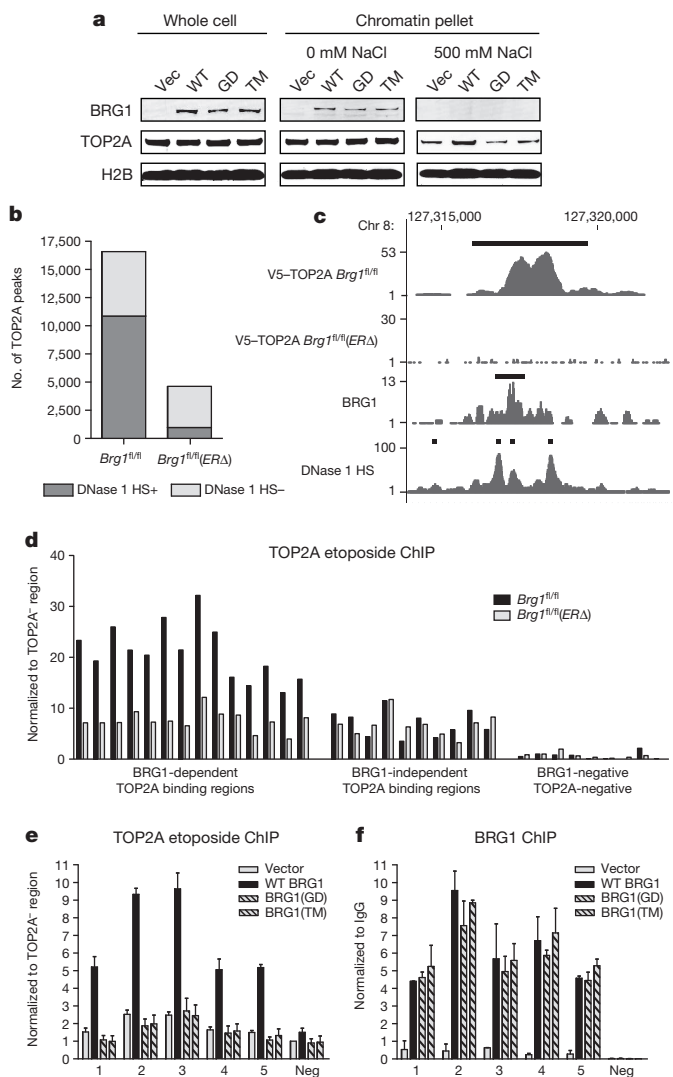


remained associated with chromatin after high salt wash in WT BRG1 cells than in BRG1(TM), BRG1(GD) and vector cells (Fig. 3a and Supplementary Fig. 5b, c). Reduced binding of TOP2A to chromatin would be expected to compromise TOP2A function and could represent an inability of TOP2A to associate with substrate DNA during decatenation.

To identify defined regions of TOP2A binding across the genome, we performed a TOP2A chromatin immunoprecipitation combined with massively parallel DNA sequencing (ChIP-seq) in *Brg1<sup>fl/fl</sup>* and *Brg1<sup>fl/fl</sup>(ERΔ)* cells. We recovered very few peaks using traditional ChIP methods, so we used etoposide, a small molecule that freezes TOP2A in a covalent complex with DNA during the enzymatic process, thereby identifying sites of active TOP2A cleavage<sup>26</sup>. We recovered 16,591 TOP2A peaks in *Brg1<sup>fl/fl</sup>* cells and 4,623 TOP2A peaks in *Brg1<sup>fl/fl</sup>(ERΔ)* cells, demonstrating the contribution of BRG1 to TOP2A binding (Fig. 3b). Almost two-thirds of the TOP2A *Brg1<sup>fl/fl</sup>* peaks are DNase-I-hypersensitive, consistent with the preference of TOP2A for nucleosome-free DNA<sup>27</sup>. An example reflecting these trends is shown in Fig. 3c. We confirmed TOP2A binding by ChIP-quantitative PCR (qPCR) at 14 BRG1-dependent and 10 BRG1-independent sites in *Brg1<sup>fl/fl</sup>* and *Brg1<sup>fl/fl</sup>(ERΔ)* cells (Fig. 3d). In addition, we determined that TOP2A binding is mitigated in BRG1(TM) and BRG1(GD) mutant *Brg1<sup>fl/fl</sup>(ERΔ)* cells at BRG1-dependent sites (Fig. 3e). This is not the result of reduced binding of the BRG1 mutants to chromatin, as BRG1(TM) and BRG1(GD) bind similarly to WT BRG1 at these sites (Fig. 3f). Given that the BRG1(TM) and BRG1(GD) mutants display reduced ATPase activity, these data implicate a role for the ATP-dependent accessibility activity of BAF complexes in TOP2A binding and function across the genome, a function previously identified for yeast protein Snf5 in transcription<sup>28</sup>.

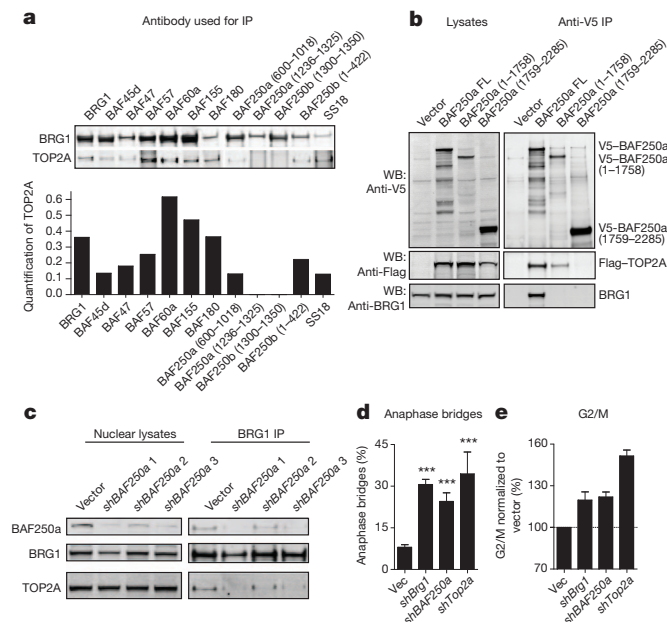
Owing to the dedicated nature of subunits within BAF complexes, TOP2A could be interacting with any BAF subunit. Indeed, we precipitated TOP2A with antibodies to several dedicated subunits as determined by glycerol gradient centrifugation analysis (Fig. 4a and Supplementary Fig. 6a). Quantification of the precipitated TOP2A revealed that little TOP2A was recovered after immunoprecipitation with antibodies raised against BAF250a (amino acids 1236–1325) and BAF250b (also known as ARID1B; amino acids 1300–1350), whereas other antibodies immunoprecipitated TOP2A well (Fig. 4a). We reasoned that the BAF250a/b antibody might disrupt the interaction between TOP2A and the BAF complex if TOP2A bound directly to BAF250a/b. Indeed, TOP2A associated with full-length BAF250a and BAF250a (amino acids 1–1758), but not BAF250a (amino acids 1759–2285) in a heterologous expression system (Fig. 4b). This interaction is independent of BRG1 because we were unable to detect BRG1 in co-precipitates of BAF250a (1–1758) and TOP2A. Furthermore, the association between TOP2A and BRG1 was lost upon knockdown of *BAF250a*, with the most severe knockdown resulting in the most severe loss of association (Fig. 4c and Supplementary Fig. 6b). To determine whether the interaction between TOP2A and BAF250a was physiologically relevant, we knocked down *BAF250a* in MEFs and observed frequencies of anaphase bridges and G2/M delay similar to knockdown of *Brg1* or *Top2a* (Fig. 4d, e and Supplementary Fig. 6c, d). These data indicate that TOP2A associates with BRG1 through a direct interaction with BAF250a.

Our studies point to a new role for ATP-dependent chromatin remodelling in decatenating DNA. Reduced decatenation *in vivo* is revealed by the frequency of anaphase bridges and an increase in the number of cells in G2/M upon deletion of BRG1 or expression of the tumour-associated T910M and G1232D BRG1 mutants (Figs 1a, d and 2c, d). Although mitotic defects have been noted in cells lacking BRG1, the cause of these defects was unclear<sup>29</sup>. In addition to BRG1, loss of BAF250a also results in decatenation defects (Fig. 4d, e), which could reflect the high incidence of mutations in BRG1 and BAF250a in human tumours<sup>1,23</sup>. Our *in vivo* observations are reinforced by the requirement of BAF for TOP2A binding at DNase-I-hypersensitive BRG1 binding sites (Fig. 3b–e).



**Figure 3 | BRG1 facilitates the binding of TOP2A to chromatin *in vivo* through ATPase-dependent chromatin-remodelling activity.** **a**, Chromatin pellets isolated from nuclei of WT BRG1-, BRG1(GD)-, BRG1(TM)- and vector-expressing *Brg1<sup>fl/fl</sup>(ERΔ)* ESCs lysed in  $\pm$  500 mM NaCl. **b**, The number of DNase-I-hypersensitive (HS) TOP2A peaks of the total number of TOP2A peaks from TOP2A ChIP-seqs in *Brg1<sup>fl/fl</sup>* and *Brg1<sup>fl/fl</sup>(ERΔ)* ESCs. **c**, Representative ChIP-seq tracks for TOP2A (in *Brg1<sup>fl/fl</sup>* and *Brg1<sup>fl/fl</sup>(ERΔ)* ESCs), BRG1 and DNase I hypersensitivity. **d**, TOP2A ChIP-qPCR confirmation from *Brg1<sup>fl/fl</sup>* and *Brg1<sup>fl/fl</sup>(ERΔ)* ESCs. **e**, TOP2A ChIP-qPCR confirmation from WT BRG1-, BRG1(GD)-, BRG1(TM)- and vector control-expressing *Brg1<sup>fl/fl</sup>(ERΔ)* ESCs. **f**, BRG1 ChIP-qPCRs from WT BRG1-, BRG1(GD)-, BRG1(TM)- and vector control-expressing *Brg1<sup>fl/fl</sup>(ERΔ)* ESCs. Data in **e** and **f** represent means of triplicate experiments  $\pm$  s.e.m.

The dependence of TOP2A on BAF function offers a possible explanation for the frequency with which BAF subunit mutations are detected in screens for driving mutations in human cancers. Anaphase bridges are often forcibly severed during cytokinesis<sup>30</sup>, resulting in partial or complete chromosome gains or losses as well as polyploidy if the cell fails to undergo mitosis<sup>12,17</sup>. At present, the number of BRG1-mutant medulloblastomas analysed for ploidy status is insufficient to determine whether BRG1 mutation results in aneuploidy in human tumours. In the case of medulloblastoma, mutations in BRG1 are often accompanied by activating mutations in the WNT signalling pathways and/or MYC amplification<sup>23</sup>. Further studies highlighting these pairings will help define the contribution of reduced TOP2A function as a result of BRG1 mutation to tumorigenesis.



**Figure 4 | TOP2A associates with the BAF complex through BAF250a.**

**a**, Immunoprecipitates from ESC nuclear lysates. Quantification of the precipitated TOP2A is shown. **b**, V5 was precipitated from HEK293T cells that had been transfected with Flag–TOP2A and either vector, V5-tagged full-length BAF250a (BAF250a FL) or V5-tagged BAF250a fragments. Lysates and anti-V5 precipitates were blotted for anti-V5, anti-Flag and anti-BRG1. **c**, BRG1 was immunoprecipitated from ESCs after BAF250a knockdown. **d**, MEFs with knockdown of *Brg1*, *BAF250a* or *Top2a*. Anaphase-bridge frequency is calculated for seven experiments  $\pm$  s.e.m. Significance was calculated relative to vector control cells where \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.0001$ . **e**, Cell cycle analysis of MEFs from **d**. Data represent the mean of the percentage of G2/M cells normalized to vector control from four experiments  $\pm$  s.e.m.

## METHODS SUMMARY

*Brg1* deletion from *Brg1*<sup>fl/fl</sup> *creER* ESCs and MEFs was performed as previously described<sup>4</sup>. Lentiviruses were produced in HEK293T cells using polyethylenimine transfection. Cells were synchronized using double thymidine block. Cell cycle analysis was performed according to manufacturer instructions (BD Biosciences). TOP2A ChIP-seq was performed following etoposide fixation<sup>26</sup>. qPCR, immunofluorescence, immunoprecipitation and western blotting were done using standard protocols. The chromatin fraction from nuclei incubated in varying concentrations of NaCl was analysed by western blot.

**Full Methods** and any associated references are available in the online version of the paper.

Received 4 June 2012; accepted 3 April 2013.

Published online 22 May 2013.

- Kadoch, C. *et al.* Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. *Nature Genet.* <http://dx.doi.org/10.1038/ng.2628> (5 May 2013).
- You, J. S. & Jones, P. A. Cancer genetics and epigenetics: two sides of the same coin? *Cancer Cell* **22**, 9–20 (2012).
- Clapier, C. R. & Cairns, B. R. The biology of chromatin remodeling complexes. *Annu. Rev. Biochem.* **78**, 273–304 (2009).
- Ho, L. *et al.* esBAF facilitates pluripotency by conditioning the genome for LIF/STAT3 signalling and by regulating polycomb function. *Nature Cell Biol.* **13**, 903–913 (2011).
- Wilson, B. G. *et al.* Epigenetic antagonism between polycomb and SWI/SNF complexes during oncogenic transformation. *Cancer Cell* **18**, 316–328 (2010).
- Parsons, D. W. *et al.* The genetic landscape of the childhood cancer medulloblastoma. *Science* **331**, 435–439 (2011).
- Pugh, T. J. *et al.* Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations within a broad landscape of genetic heterogeneity. *Nature* **488**, 106–110 (2012).

- Jones, D. T. W. *et al.* Dissecting the genomic complexity underlying medulloblastoma. *Nature* **488**, 100–105 (2012).
- Robinson, G. *et al.* Novel mutations target distinct subgroups of medulloblastoma. *Nature* **488**, 43–48 (2012).
- Love, C. *et al.* The genetic landscape of mutations in Burkitt lymphoma. *Nature Genet.* **44**, 1321–1325 (2012).
- Richter, J. *et al.* Recurrent mutation of the ID3 gene in Burkitt lymphoma identified by integrated genome, exome and transcriptome sequencing. *Nature Genet.* **44**, 1316–1320 (2012).
- Carpenter, A. J. & Porter, A. C. Construction, characterization, and complementation of a conditional-lethal DNA topoisomerase II $\alpha$  mutant human cell line. *Mol. Biol. Cell* **15**, 5700–5711 (2004).
- Lou, Z., Minter-Dykhouse, K. & Chen, J. BRCA1 participates in DNA decatenation. *Nature Struct. Mol. Biol.* **12**, 589–593 (2005).
- Dawlaty, M. M. *et al.* Resolution of sister centromeres requires RanBP2-mediated SUMOylation of topoisomerase II $\alpha$ . *Cell* **133**, 103–115 (2008).
- Ramamoorthy, M. *et al.* RECQL5 cooperates with Topoisomerase II alpha in DNA decatenation and cell cycle progression. *Nucleic Acids Res.* **40**, 1621–1635 (2012).
- Ho, L. *et al.* An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency. *Proc. Natl Acad. Sci. USA* **106**, 5181–5186 (2009).
- Johnson, M., Phua, H. H., Bennett, S. C., Spence, J. M. & Farr, C. J. Studying vertebrate topoisomerase 2 function using a conditional knockdown system in DT40 cells. *Nucleic Acids Res.* **37**, e98 (2009).
- Downes, C. S. *et al.* A topoisomerase II-dependent G2 cycle checkpoint in mammalian cells. *Nature* **372**, 467–470 (1994).
- Luo, K., Yuan, J., Chen, J. & Lou, Z. Topoisomerase II $\alpha$  controls the decatenation checkpoint. *Nature Cell Biol.* **11**, 204–210 (2009).
- Sakaguchi, A. & Kikuchi, A. Functional compatibility between isoform  $\alpha$  and  $\beta$  of type II DNA topoisomerase. *J. Cell Sci.* **117**, 1047–1054 (2004).
- Khavari, P. A., Peterson, C. L., Tamkun, J. W., Mendel, D. B. & Crabtree, G. R. BRG1 contains a conserved domain of the SWI2/SNF2 family necessary for normal mitotic growth and transcription. *Nature* **366**, 170–174 (1993).
- Kool, M. *et al.* Molecular subgroups of medulloblastoma: an international meta-analysis of transcriptome, genetic aberrations, and clinical data of WNT, SHH, Group 3, and Group 4 medulloblastomas. *Acta Neuropathol.* **123**, 473–484 (2012).
- Northcott, P. A. *et al.* Medulloblastomics: the end of the beginning. *Nature Rev. Cancer* **12**, 818–834 (2012).
- Ho, L. *et al.* An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. *Proc. Natl Acad. Sci. USA* **106**, 5187–5191 (2009).
- Stros, M., Bacikova, A., Polanska, E., Stokrova, J. & Strauss, F. HMGB1 interacts with human topoisomerase II $\alpha$  and stimulates its catalytic activity. *Nucleic Acids Res.* **35**, 5001–5013 (2007).
- Sano, K., Miyaji-Yamaguchi, M., Tsutsui, K. M. & Tsutsui, K. Topoisomerase II $\beta$  activates a subset of neuronal genes that are repressed in AT-rich genomic environment. *PLoS ONE* **3**, e4103 (2008).
- Capranico, G., Jaxel, C., Roberge, M., Kohn, K. W. & Pommier, Y. Nucleosome positioning as a critical determinant for the DNA cleavage sites of mammalian DNA topoisomerase II in reconstituted simian virus 40 chromatin. *Nucleic Acids Res.* **18**, 4553–4559 (1990).
- Sperling, A. S., Jeong, K. S., Kitada, T. & Grunstein, M. Topoisomerase II binds nucleosome-free DNA and acts redundantly with topoisomerase I to enhance recruitment of RNA Pol II in budding yeast. *Proc. Natl Acad. Sci. USA* **108**, 12693–12698 (2011).
- Bourgo, R. J. *et al.* SWI/SNF deficiency results in aberrant chromatin organization, mitotic failure, and diminished proliferative capacity. *Mol. Biol. Cell* **20**, 3192–3199 (2009).
- Janssen, A., van der Burg, M., Szuhai, K., Kops, G. J. & Medema, R. H. Chromosome segregation errors as a cause of DNA damage and structural chromosome aberrations. *Science* **333**, 1895–1898 (2011).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We would like to acknowledge the DNA Sequencing Core facility of the National Heart, Lung, and Blood Institute (NHLBI) for sequencing the ChIP-seq libraries. This work was supported by the NIH (G.R.C.), The American Cancer Society (E.C.D.), The Helen Hay Whitney Foundation (D.C.H.) and the Division of Intramural Research program of NHLBI (K.Z.). G.R.C. is an investigator at the Howard Hughes Medical Institute.

**Author Contributions** E.C.D., D.C.H. and G.R.C. designed the experiments, and E.C.D. and D.C.H. performed the experiments. E.M. performed bioinformatic analysis. A.K., M.K., S.P. and Y.-J.C. contributed human tumour samples. K.Z. and K.C. performed sequencing of the TOP2A ChIP. E.C.D., D.C.H. and G.R.C. wrote the manuscript.

**Author Information** The TOP2A ChIP-seq is deposited in the Gene Expression Omnibus (GEO) under accession number GSE45625. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.R.C. (crabtree@stanford.edu).

## METHODS

*Brg1* deletion from *Brg1<sup>fl/fl</sup> creER* ESCs and MEFs was performed as previously described<sup>4</sup>. Lentiviruses were produced in HEK293T cells using polyethylenimine (PEI) transfection. Cells were synchronized using double thymidine block. Cell cycle analysis was performed according to manufacturer instructions (BD Biosciences). TOP2A ChIP-seq was performed following etoposide fixation<sup>26</sup>. qPCR, immunofluorescence, immunoprecipitation and western blotting were done using standard protocols. The chromatin fraction from nuclei in varying concentrations of NaCl was analysed by western blot.

**Immunofluorescence.** To quantify anaphase bridges, cells were fixed with 4% paraformaldehyde for 20 min, washed and stained with 4',6-diamidino-2-phenylindole (DAPI; Sigma). The number of anaphases/telophases with bridges over the total number of anaphases (between 56 and 187 total anaphases per 25-mm slide) was recorded from each slide for more than four independent experiments. *Brg1<sup>fl/fl</sup>* and *Brg1<sup>fl/fl</sup>(ERA)* ESCs were visualized with DAPI 72 h after tamoxifen treatment. Wild-type MEFs were infected with lentiviruses containing short hairpin RNA (shRNA) to *Brg1*, *BAF250a* or *Top2a* and analysed 48–96 h after infection.

To stain for TOP2A and centromeres/microtubules, cells were blocked with 5% BSA/1% goat serum in phosphate buffered saline with 0.1% Tween-20 (PBST) for 1 h following fixation and incubated with anti-centromere (Antibodies Inc.) or anti- $\gamma$ -tubulin (Sigma) and anti-TOP2A (Santa Cruz) for 2 h. After several washes, anti-human Alexa Fluor 488 and anti-rabbit Alexa Fluor 568 were added for 1 h. The cells were then stained with DAPI for 10 min and washed 3  $\times$  PBS for 10 min each. The coverslips were mounted on slides with Vectashield Hard Mount (Vector Labs).

To quantify anaphase bridges from paraffin-embedded human tumour samples, slides were incubated 2  $\times$  25 min in xylenes, then rehydrated in 100% ethanol, then 95% ethanol, then water, for 2 min each. The slides were boiled in citrate buffer (pH 6.0) (Vector Labs) for 20 min and washed 2  $\times$  5 min in PBS-Tween. The slides were then stained with DAPI for 10 min and washed 3  $\times$  5 min with PBS before mounting with Vectashield Hard Mount.

**Cell synchronization.** ESCs were incubated with 2 mM thymidine for 7–8 h, released into fresh media for 7 h, and then incubated with thymidine again for 7 h. The cells were washed several times with PBS, released into fresh media, and collected at time points thereafter.

**Cell cycle analysis.** The cell cycle analysis was performed using BD Biosciences BrdU-FITC FACS kit. ESCs were incubated with BrdU for 1 h and MEFs were incubated with BrdU for 4 h. *Brg1<sup>fl/fl</sup>* and *Brg1<sup>fl/fl</sup>(ERA)* ESCs were analysed 72 h after tamoxifen treatment. Caffeine was added to media 2 h before BrdU incubation. To determine the per cent of cells in G2/M, DNA was stained with 7-AAD and analysed by FACS.

**H3(S10)P cell cycle analysis.** *Brg1<sup>fl/fl</sup>* ESCs were infected with interference-RNA-resistant wild-type human TOP2A or TOP2A(S1524A) and shRNAs to mouse *Top2a*. Cells were stained with anti-H3(S10)P and analysed by flow cytometry 72 h after treatment with or without tamoxifen.

**Metaphase spread preparation.** MEFs were grown to 85% confluence and incubated for 4 h with colcemid. Cells were collected and swelled by drop-wise addition of 1:1 0.4% KCl/0.4% sodium citrate for 7 min at 37 °C. Cells were then fixed by dropwise addition of 3:1 methanol/acetic acid for 20 min, spun down and fixed for another 30 min. Metaphases were dropped onto slides, dried on wet paper towels and stained with DAPI for visualization. Chromosomes were then measured and counted using ImageJ software. To analyse polyploidy, only cells with greater than 35 chromosomes were counted to eliminate artefacts due to partial spreads.

**Gene expression profiling and analysis.** RNA was isolated using TRIzol (Invitrogen) and reverse transcribed into complementary DNA using SuperScript III reverse transcriptase (Invitrogen). qPCR was performed on the StepOnePlus (ABI) machine using FastStart Universal SYBR Green Master with ROX (Roche).

**Immunoprecipitation.** Nuclei were isolated from cells with buffer A (25 mM HEPES, pH 7.6, 5 mM MgCl<sub>2</sub>, 25 mM KCl, 0.05 mM EDTA, 10% glycerol, 0.1% NP-40) and lysed for 30 min in immunoprecipitation buffer (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.1% NP-40). The chromatin was removed using centrifugation and the lysates were pre-cleared with 20  $\mu$ l Protein A or Protein G Dynabeads for 30 min. The protein concentration was quantitated using the bicinchoninic acid (BCA) assay (Pierce) and adjusted to a final volume of 200  $\mu$ l at a final concentration of 1.5 mg ml<sup>-1</sup> with immunoprecipitation buffer. Each immunoprecipitation was incubated with 3  $\mu$ g of anti-BRG1 (Santa Cruz), anti-TOP2A (Abcam), anti-BAF45d (Crabtree laboratory), anti-BAF47 (Santa Cruz), anti-BAF57 (Bethyl), anti-BAF155 (Crabtree laboratory), anti-BAF60a (BD Transduction Laboratories), anti-BAF250a (Santa Cruz), anti-BAF180 (Bethyl), anti-BAF250b (Santa Cruz, Bethyl), anti-SS18 (Santa Cruz) anti-BAF200 (Santa Cruz) or anti-IgG (Santa Cruz) overnight at 4 °C and then for 2 h with 20  $\mu$ l

Protein A/G Dynabeads. The beads were washed four times with 1 ml immunoprecipitate buffer and re-suspended in 10  $\mu$ l gel loading buffer (4 $\times$  lithium dodecyl sulfate buffer; Invitrogen).

**Glycerol gradient centrifugation analysis.** ESCs were lysed in buffer A (10 mM HEPES, pH 7.6), 25 mM KCl, 1 mM EDTA, 10% glycerol, 1 mM DTT and protease inhibitors (complete mini tablets (Roche) supplemented with 1 mM phenylmethylsulphonyl fluoride) on ice. Nuclei were sedimented by centrifugation (1,000g), re-suspended in buffer C (10 mM HEPES, pH 7.6), 3 mM MgCl<sub>2</sub>, 100 mM KCl, 0.1 mM EDTA, 10% glycerol, 1 mM dithiothreitol (DTT) and protease inhibitors, and lysed by the addition of ammonium sulphate to a final concentration of 0.3 M. Soluble nuclear proteins were separated by insoluble chromatin fraction by ultracentrifugation (100,000g) and precipitated with 0.3 mg ml<sup>-1</sup> ammonium sulphate for 20 min on ice. Protein precipitate was isolated by ultracentrifugation (100,000g) and re-suspended in HEMG-0 buffer (25 mM HEPES, pH 7.9, 0.1 mM EDTA, 12.5 mM MgCl<sub>2</sub>, 100 mM KCl) for glycerol gradient analyses. 800  $\mu$ g of protein was overlaid on to a 10-ml 10–30% glycerol (in HEMG buffer) gradient prepared in a 14  $\times$  89 mm polyallomer centrifuge tube (Beckman). Tubes were placed in a SW-40 swing bucket rotor and centrifuged at 4 °C for 16 h at 40,000 r.p.m. 0.5-ml fractions were collected and used in gel electrophoresis and subsequent western blotting analyses.

**Western blots.** Nuclei were isolated from cells with buffer A (25 mM HEPES, pH 7.6, 5 mM MgCl<sub>2</sub>, 25 mM KCl, 0.05 mM EDTA, 10% glycerol, 0.1% NP-40) and lysed for 30 min in RIPA buffer (10 mM Tris, pH 7.4, 150 mM NaCl, 0.1% SDS, 0.5% sodium deoxycholate, 1% Triton X-100, 1 mM EDTA). Chromatin was either spun out or samples were sonicated before BCA analysis. Equal amounts of protein were boiled in gel loading buffer and loaded onto 4–10% BisTris NuPage gels. After transfer, blots were blocked in 5% BSA and incubated with anti-BRG1, anti-TOP2A, anti-phosphotyrosine (Millipore), anti-phosphoserine (Millipore) or anti-ubiquitin (Santa Cruz). Proteins were detected using the LI-COR detection system or enhanced chemiluminescence (ECL)/autoradiography for SUMO-TOP2A.

**Chromatin association assay.** Nuclei were isolated using buffer A and re-suspended in 20 mM Tris-HCl, pH 7.6, 3 mM EDTA at 60 million cells per ml. Samples of 25  $\mu$ l were aliquotted into tubes and NaCl concentrations were adjusted to a final volume of 50  $\mu$ l. Samples were gently mixed and incubated on ice for 20 min and centrifuged at high speed for 20 min to isolate chromatin. The lysate was removed and the chromatin pellet was re-suspended in 120  $\mu$ l gel loading dye. The pellet was solubilized using sonication and the association of TOP2A to chromatin was analysed using western blotting.

**ATPase assay.** ATPase assay was adapted from the literature<sup>31</sup>. Immunoprecipitations were performed as described above with anti-BRG1 antibody. Immunoprecipitates were washed a final time with 10 mM Tris-HCl, pH 7.5, 50 mM NaCl, 5 mM MgCl<sub>2</sub>, 1 mM DTT and re-suspended in 20  $\mu$ l assay buffer (10 mM Tris-HCl, pH 7.5, 50 mM NaCl, 5 mM MgCl<sub>2</sub>, 20% glycerol, 1 mg ml<sup>-1</sup> BSA, 0.5 mM ATP, 20 nM plasmid DNA, 50  $\mu$ Ci ml<sup>-1</sup>  $\gamma$ -<sup>32</sup>P ATP, 1 mM DTT, protease inhibitors). The reaction was agitated at 37 °C for 1 h (or when approximately 50% of ATP was converted to inorganic phosphate). Reaction mixture (0.5  $\mu$ l) was spotted onto PEI cellulose plates and thin layer chromatography was performed in 0.5 M LiCl and 1 M formic acid. The plates were dried and imaged using phosphorimaging. The enzymatic activity was quantitated as a ratio of product ([<sup>32</sup>P]P<sub>i</sub>) to starting material ([ $\gamma$ -<sup>32</sup>P]ATP). Values were normalized to the activity of wild-type BRG1 (100%) and vector control (0%) cells.

**Chromatin immunoprecipitation.** For the BRG1 ChIP, 40 million ESCs were fixed for 12 min in 1% formaldehyde at room temperature (25 °C). Nuclei were sonicated in 1 ml ChIP lysis buffer (50 mM HEPES, pH 7.5, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.1% SDS) to yield fragments between 200–500 base pairs. Five-hundred microlitres of lysate was incubated with 5  $\mu$ g of anti-BRG1 (Crabtree laboratory) or 5  $\mu$ g anti-rabbit IgG and rotated overnight at 4 °C and then for 2 h with 20  $\mu$ l Protein A/G Dynabeads. After five washes with ChIP lysis buffer and one wash in 10 mM Tris-HCl, pH 8.0, 1 mM EDTA (TE), DNA was eluted by boiling in 10% Chelex slurry.

The etoposide ChIP of TOP2A was adapted from the literature<sup>26</sup>. Specifically, 20 million ESCs were treated with 100  $\mu$ M etoposide for 10 min. Cells were washed once with PBS and lysed with 1 ml of a buffer containing 1% Sarkosyl, 10 mM Tris-HCl, pH 7.5, 10 mM EDTA and protease inhibitor. A solution of 7 M CsCl (7 M) was added to a final concentration of 0.5 M and the lysate was sonicated to yield fragments between 200–500 base pairs. ChIP buffer (300  $\mu$ l) was added to 300  $\mu$ l of lysate for a final concentration of 50 mM HEPES, pH 7.5, 300 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% DOC, and 0.1% SDS and 3  $\mu$ g anti-TOP2A pre-bound to 20  $\mu$ l Protein G Dynabeads was added. The lysate was rotated overnight at 4 °C and washed four times with ChIP lysis buffer, one time with LiCl buffer (10 mM Tris, pH 8.0, 0.25 M LiCl, 0.5% NP-40, 0.5% DOC, 1 mM EDTA) and one time with TE. The DNA was eluted with 300  $\mu$ l of 1% SDS, 0.1 M NaHCO<sub>3</sub> for



20 min and removed from the beads. The solution was adjusted to 200 mM NaCl, 10 mM EDTA, 40 mM Tris, pH 6.5, and 0.2 mg ml<sup>-1</sup> RNase A was added for 30 min at 37 °C. Proteinase K was added to 0.03 mg ml<sup>-1</sup> and digested overnight at 55 °C. The DNA was extracted with phenol/chloroform and precipitated with ethanol for analysis by qPCR. Primers used for ChIP-qPCR are available upon request.

**ChIP-seq and analysis.** The library preparation and sequencing was performed as previously described<sup>32</sup>. Raw ChIP-seq reads were mapped to the *Mus musculus* genome (build mm9/NCBI37) using the short-read aligner Bowtie (version 0.12.7)<sup>33</sup>. Peaks were then called using Model-based analysis of ChIP-seq (MACS) (version 1.4.1)<sup>34</sup>. Further analysis was aided by the Bedtools suite (version 2.16.2)<sup>35</sup>. Genome annotations were acquired from the UCSC Genome Browser (<http://genome.ucsc.edu/>)<sup>36,37</sup>. We also uploaded our data to the genome browser, which was used to produce screenshots of chromatin binding/modification profiles at individual loci.

**Topoisomerase activity assay.** Reactions contain: 150 ng kinetoplast DNA (Topogen), 50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 10 mM MgCl<sub>2</sub>, 2 mM ATP, a standard TOP2A immunoprecipitate or varying amounts of recombinant TOP2A (Topogen).

**Lentiviral infection.** HEK293T cells were transfected with lentiviruses containing vector alone, wild-type BRG1, BRG1 point mutants, wild-type human TOP2A or human TOP2A(S1524A), or with vectors containing shRNA to *Brg1* (clone ID

TRCN0000071386), *Arid1a* (clone ID TRCN0000071395, Origene) or *Top2a* (clone ID V2LMM\_11295). After 48 h, supernatants were collected and centrifuged at 20,000 r.p.m. for 2 h. Viral pellets were re-suspended in PBS and used to infect ESCs in suspension or MEFs by spinfection. Cells were selected with puromycin and collected 48–96 h after infection for analysis.

31. Bultman, S. J., Gebuhr, T. C. & Magnuson, T. A. Brg1 mutation that uncouples ATPase activity from chromatin remodeling reveals an essential role for SWI/SNF-related complexes in  $\beta$ -globin expression and erythroid development. *Genes Dev.* **19**, 2849–2861 (2005).
32. Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837 (2007).
33. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
34. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
35. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
36. Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
37. Meyer, L. R. *et al.* The UCSC Genome Browser database: extensions and updates 2013. *Nucleic Acids Res.* **41**, D64–D69 (2012).

# Specialized filopodia direct long-range transport of SHH during vertebrate tissue patterning

Timothy A. Sanders<sup>1,2\*</sup>, Esther Llagostera<sup>1,3,4\*</sup> & Maria Barna<sup>1,3,4</sup>

The ability of signalling proteins to traverse tissues containing tightly packed cells is of fundamental importance for cell specification and tissue development; however, how this is achieved at a cellular level remains poorly understood<sup>1</sup>. For more than a century, the vertebrate limb bud has served as a model for studying cell signalling during embryonic development<sup>2</sup>. Here we optimize single-cell real-time imaging to delineate the cellular mechanisms for how signalling proteins, such as sonic hedgehog (SHH), that possess membrane-bound covalent lipid modifications traverse long distances within the vertebrate limb bud *in vivo*. By directly imaging SHH ligand production under native regulatory control in chick (*Gallus gallus*) embryos, our findings show that SHH is unexpectedly produced in the form of a particle that remains associated with the cell via long cytoplasmic extensions that span several cell diameters. We show that these cellular extensions are a specialized class of actin-based filopodia with novel cytoskeletal features that have not been previously described. Notably, particles containing SHH travel along these extensions with a net anterograde movement within the field of SHH cell signalling. We further show that in SHH-responding cells, specific subsets of SHH co-receptors, including cell adhesion molecule downregulated by oncogenes (CDO) and brother of CDO (BOC), actively distribute and co-localize in specific micro-domains within filopodial extensions, far from the cell body. Stabilized interactions are formed between filopodia containing SHH ligand and those containing co-receptors over a long range. These results suggest that contact-mediated release propagated by specialized filopodia contributes to the delivery of SHH at a distance. Together, these studies identify an important mode of communication between cells that considerably extends our understanding of ligand movement and reception during vertebrate tissue patterning.

The regulated movement of key signalling proteins within tissues is a central feature of metazoan development that remains poorly understood at the cellular level<sup>1</sup>. Several mechanisms have been proposed to distribute signalling molecules, including free diffusion, transcytosis and directed transport of signalling receptors via filopodia, including those termed cytonemes within invertebrate embryos<sup>1,3–5</sup>. An example of a key signalling protein is SHH, which is instrumental in patterning the early embryo<sup>6</sup>. During limb development, SHH is produced by the zone of polarizing activity (ZPA), a small group of mesenchymal cells at the posterior margin of the limb bud, and it acts over a long range to specify the number and identity of digits produced<sup>7</sup>. How tight control in SHH distribution across several cell diameters is established remains poorly understood. Here we combine unique genetic and live-cell imaging approaches to investigate the mechanisms of long-range cell signalling underlying tissue patterning during vertebrate embryonic development.

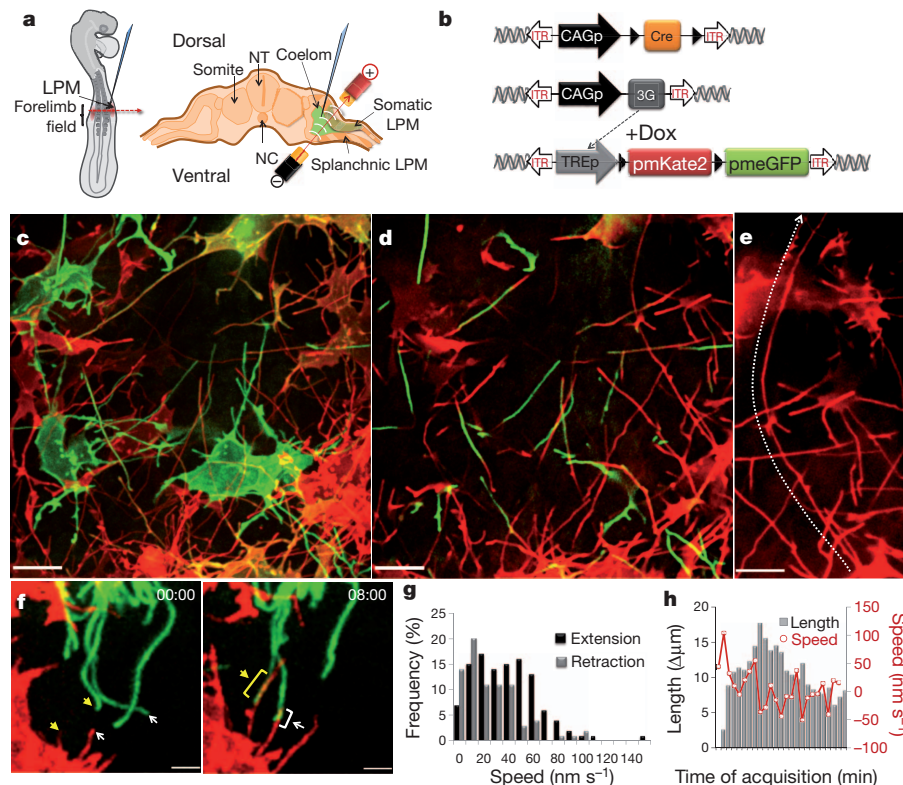
To image living chick embryos under the regulatory control of specific spatial and temporal regulatory elements, we developed a piggyBac transposon-mediated stable integration approach<sup>8</sup>. This genetically tractable expression system, coupled with optimized confocal microscopy,

enabled imaging of embryos (Methods) at single-cell and subcellular resolution (Fig. 1a and Supplementary Fig. 1). At first, we expressed membrane-tethered fluorescent proteins that illuminate individual mesenchymal cells within the developing limb bud in a mosaic fashion (Fig. 1b). Through this approach we uncovered an unexpected, intricate network of thin cellular extensions present on these cells spanning several cell diameters, which orient in many directions from the cell body along the anterior–posterior, proximal–distal and dorsal–ventral axis (Fig. 1c–e and Supplementary Fig. 2). All mesenchymal cells possess several cellular extensions up to 150  $\mu\text{m}$  long, with an average length of  $34.27 \pm 9.6 \mu\text{m}$  (mean  $\pm$  s.e.m.) ( $n = 24$ ), revealing a surprising morphology to these cells (Fig. 1c–e and Supplementary Video 1). These extensions are remarkably fine, approximately 200 nm in diameter, at the resolving limit of conventional microscopy and can be labelled with membrane-bound (Fig. 1) but not cytoplasmic (Supplementary Fig. 3) fluorescent proteins. They are capable of elongating, at a maximum rate of  $150 \text{ nm s}^{-1}$ , retracting, and traversing the complex three-dimensional extracellular matrix of the limb bud (Fig. 1g, h, Supplementary Fig. 2 and Supplementary Videos 2 and 3). Importantly, despite their dynamic nature, they form highly stabilized long-range interactions between cells, thereby revealing a new and complex landscape of cell–cell interactions mediated through cytoplasmic extensions within embryonic tissues (Fig. 1f, Supplementary Fig. 2 and Supplementary Videos 4 and 5). These extensions are not amenable to conventional fixation, which severely disrupts their structure and may have precluded their previous identification (Supplementary Fig. 5).

Actin-associated markers—including the high-affinity F-actin probe utrophin calponin homology domain (UCHD) fused to enhanced green fluorescent protein (UCHD–eGFP) and moesin–eGFP—decorate the entire length of cytoplasmic extensions, revealing that these structures are actin-based filopodia (Fig. 2a and Supplementary Fig. 4a). Unlike actin markers, tubulin cytoskeleton markers, such as microtubule-associated proteins TAU (also known as MAPT) and EB3, only label the proximal base of a subset of mesenchymal filopodia (data not shown). We next examined the localization of a plus-ended actin motor, myosin-X<sup>9</sup>. Notably, myosin-X–eGFP moves to the distal tips of the filopodia, where it accumulates, thereby revealing that actin motors can move along these structures (Fig. 2b and data not shown).

Limb bud mesenchymal cytoplasmic extensions also possess distinct cytoskeletal features compared to typical filopodia, commonly characterized as actin-based linear extensions of the cell membrane, with limited lengths of up to 10  $\mu\text{m}$  (ref. 10). For example, LifeAct is a highly specific marker of filopodia in eukaryotic cells, but it unexpectedly labels only the proximal base of limb mesenchyme cytoplasmic extensions (Fig. 2c) and not their distal tips, reflecting a distinguishing feature. This is consistent with the fact that LifeAct does not label certain forms of highly modified actin, or actin that is highly coated with binding proteins<sup>11</sup>. Additional specific features are also evident in the actin depolymerization factor cofilin, which has mainly been implicated in extending lamellipodial protrusions<sup>12</sup>. Notably, cofilin–eGFP rapidly accumulates to the

<sup>1</sup>Department of Biochemistry and Biophysics, University of California San Francisco, San Francisco, California 94158, USA. <sup>2</sup>Department of Pediatrics, Division of Neonatology, University of California San Francisco, San Francisco, California 94158, USA. <sup>3</sup>Department of Developmental Biology, Stanford University, Stanford, California 94305, USA. <sup>4</sup>Department of Genetics, Stanford University, Stanford, California 94305, USA. \*These authors contributed equally to this work.



**Figure 1 | Mesenchymal cells of the developing limb bud possess long and highly dynamic cytoplasmic extensions.** **a**, Left, HH14 chick embryo indicating the site of DNA injection. The red dashed line indicates the cross-sectional plane. Right, microinjected DNA in the coelom is shown in green; the electrodes position is indicated (LPM). NC, notochord; NT, neural tube. **b**, Diagram of piggyBac transposon system resulting in integration of transposon inverted terminal repeat (ITR)-flanked expression cassettes by piggyBac transposase. Cre recombinase flanked by *loxP* sites (black triangles) containing membrane-palmitoylated mKate2 (a far-red fluorescent protein) or pmeGFP expressed via the tetracycline responsive element (TRE). Dox, doxycycline. CAGp denotes a ubiquitous promoter; 3G denotes an inducible transactivator protein. **c**, Confocal z-series acquired *in vivo* from an HH21 limb

bud reveals an intricate network of cellular extensions (Supplementary Video 1). **d**, Single *x-y* plane, from **c**, highlighting the network of long cytoplasmic extensions among mesenchymal cells. **e**, A representative long extension (75  $\mu\text{m}$ ) from **c**, marked by dotted line. **f**, Example of an interaction between two cytoplasmic extensions. Cytoplasmic extensions emanating from two cells that are initially separated (left panel, yellow and white arrows) then extend until they interact and overlap (right panel, yellow and white brackets) to form stabilized interactions over a long-range (see also Supplementary Video 4). Time in min:s. **g**, Speed distribution of extending (black) and retracting (grey) cytoplasmic extension velocities;  $n = 8$ . **h**, Extension dynamics. Grey bars represent net length change in micrometres. Red line represents the mean velocity ( $\text{nm s}^{-1}$ ). The *x* axis tick marks denote 1-min intervals. Scale bars, 10  $\mu\text{m}$  (**c–e**), 3  $\mu\text{m}$  (**f**).

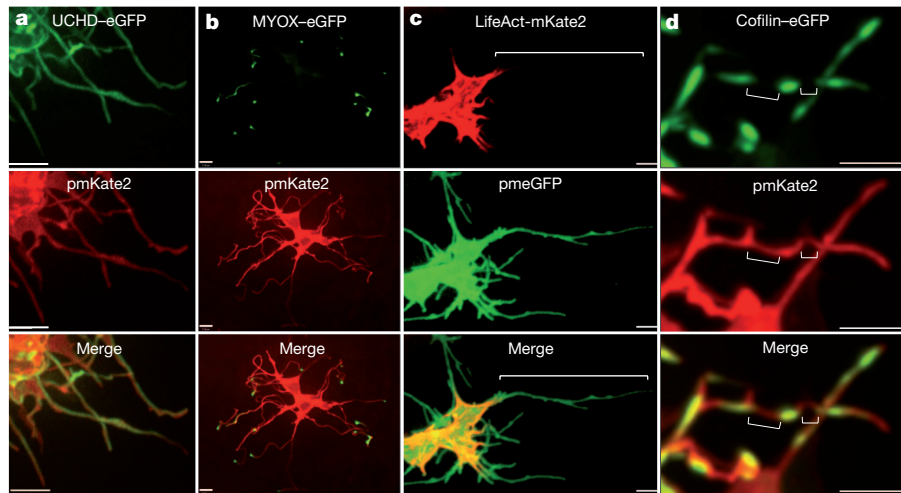
tips of limb mesenchymal filopodia, and its subsequent retraction back to the cell soma prefigures the rapid and dynamic retraction of filopodia extensions (Supplementary Fig. 4b and Supplementary Video 6). Cofilin-eGFP is also frequently localized to specific microdomains along these filopodia that are interrupted rather than labelling the entire process that may account, at least in part, for the greater dynamics of these filopodia, including movement and bends in specific sub-regions as they traverse extracellular space (Fig. 2d). Fascin, which enhances cofilin severing<sup>12</sup>, also labels filopodial extensions (Supplementary Fig. 4c). Together, these findings demonstrate that limb mesenchyme filopodia possess unique cytoskeletal features reflecting specialized properties, which include their considerable lengths, highly dynamic behaviours, and complex geometries. Our initial attempts to perturb mesenchymal filopodia formation using known molecular pathways (Methods) have proven ineffective, for example, through conditional inactivation of cell division cycle 42 (CDC42) in the limb bud (data not shown).

To determine the functional role of mesenchymal filopodia, we used genetic strategies to label specific cellular populations with membrane-bound fluorescent proteins along the anterior–posterior axis of the mouse limb bud with respect to SHH signalling. This revealed that SHH-producing cells within the limb bud ZPA extend long filopodia (Supplementary Fig. 6a and Supplementary Video 7), which can orient along the anterior–posterior axis as well as the proximal–distal axis, with a further bias towards the apical ectodermal ridge that maintains

the SHH and fibroblast growth factor (FGF) feedback loop<sup>13</sup> (Supplementary Fig. 7a, b). Moreover, mesenchymal cells within the anterior limb bud that respond to SHH also extend similar filopodia (Supplementary Fig. 6b); however, they show a bias in orientation along the anterior–posterior axis (Supplementary Fig. 7a, c). Thereby, these results demonstrate that both SHH-producing and SHH-responding cells extend specialized filopodia within the SHH signalling field.

To determine the possible role of limb mesenchyme filopodia with respect to SHH signalling, we visualized key components of the SHH signalling pathway. To image SHH ligand, we developed a tightly regulated expression system directed by the endogenous *G. gallus* SHH minimal promoter and limb-specific enhancer element (ZPA regulatory sequence (ZRS))<sup>14</sup>, coupled with doxycycline-inducible control to allow for transient gene expression in a small number of cells within the ZPA (Supplementary Fig. 8a). SHH is produced as a precursor protein that undergoes autocatalytic processing to yield an amino-terminal signalling protein (SHH<sub>p</sub>) containing cholesterol and palmitate moieties<sup>15</sup>. SHH fusion proteins with monomeric eGFP were constructed, including SHH<sub>p</sub>-eGFP that retains the cholesterol modification, and SHH<sub>N</sub>-eGFP that does not but produces a brighter and more photostable fluorescent fusion protein (Supplementary Fig. 8b and Methods). Consistent with the fact that SHH<sub>N</sub> can activate SHH signalling within the limb bud<sup>16</sup>, we find that after ectopic expression of SHH<sub>N</sub>-eGFP, patched (PTC1) a marker of SHH activation, is expanded anteriorly within the limb bud



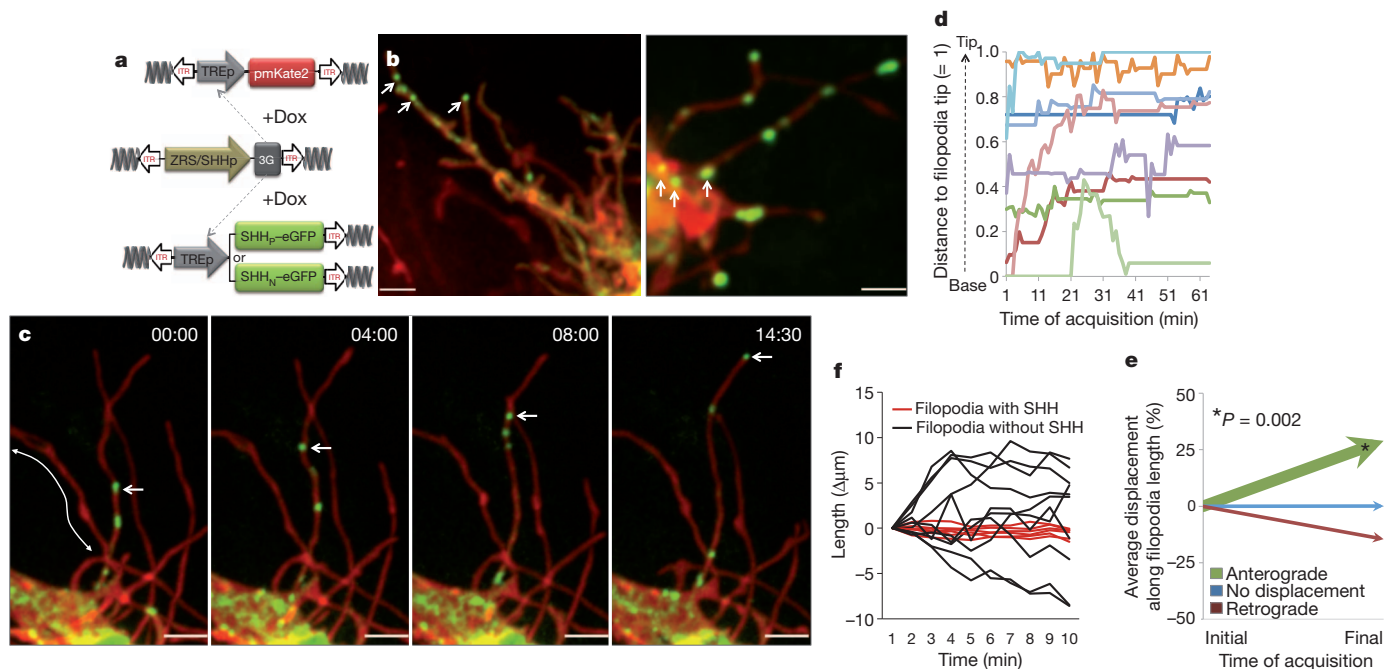


**Figure 2 | Limb mesenchymal cytoplasmic extensions are a class of specialized actin-based filopodia.** **a**, UCHD-eGFP-localization staining demonstrating that membrane-labelled pmKate2 filopodia extensions contain actin filaments. **b**, Myosin X-eGFP (MYOX-eGFP) is localized to each pmKate2-labelled filopodium and is concentrated at the distal tip.

**c**, LifeAct-mKate2 marks only the proximal aspect of pmeGFP-labelled filopodia and does not label the entire extension, shown by the bracket. **d**, Cofilin-eGFP is present in interrupted domains along the filopodia, negative regions shown with brackets. Scale bars, 3  $\mu\text{m}$  (**a**), 5  $\mu\text{m}$  (**b–d**).

(Supplementary Fig. 8e). Notably, expression of these SHH fusion proteins under doxycycline-inducible control does not perturb endogenous SHH signalling as revealed by PTC1 expression, nor limb development or skeletal patterning (Supplementary Fig. 8d).

Imaging of SHH<sub>N</sub>-eGFP under native regulatory control in the ZPA reveals that it is unexpectedly produced in the form of a particle approximately 200 nm in size (Fig. 3a, b). These particles are not observed in the extracellular space, but remain associated with the SHH-producing cell



**Figure 3 | Live-cell imaging of SHH ligand production and transport within the limb bud.** **a**, Schematic of piggyBac-mediated integration of transposon-flanked expression cassettes (ITR). The *G. gallus* SHH minimal promoter (SHHp) and ZRS element direct spatial expression in the limb ZPA of doxycycline-inducible transactivator protein (3G), which in turn allows for the temporal control of SHH<sub>N</sub>-eGFP or SHH<sub>P</sub>-eGFP and pmKate2. **b**, Left, a representative SHH-producing cell containing multiple long filopodia, with SHH<sub>P</sub>-eGFP present in discrete particles as well as in a more diffuse form localized along these extensions. Right, SHH<sub>N</sub>-eGFP is produced as a particle visualized within the cell soma (arrows) as well as along the filopodia. **c**, Representative timelapse images showing anterograde SHH<sub>N</sub>-eGFP particle movement (arrows) that accumulates at the tip of the certain filopodia but not others (indicated by double-headed arrow) (Supplementary Video 8). Time in

min:s, interval is four frames  $\text{s}^{-1}$ . **d**, Particle dynamics graph demonstrating the movement of SHH<sub>N</sub>-eGFP particles relative to the filopodium; normalized distance to filopodia base is 0, and to filopodia tip is 1. **e**, Net particle movement graph demonstrating the net vectors of particle ( $n = 38$ ) displacement represented as a percentage of the total filopodia length that particles traverse. Green denotes anterograde, blue denotes no displacement ( $<5\%$ ), and red denotes retrograde displacement. The relative thickness of each vector reflects the percentage of particles within each category. There is a statistically significant net anterograde movement of SHH particles away from the cell soma;  $P < 0.002$ . **f**, SHH-containing filopodia (red lines) are statistically more stabilized than filopodia without SHH (black lines);  $P < 0.001$ ,  $n = 200$  time points. All scale bars, 3  $\mu\text{m}$ .

via long filopodial extensions. Imaging of SHH<sub>N</sub>-eGFP also reveals that cholesterol-modified SHH is similarly produced as a particle associated with filopodial extensions, and also displays more uniform localization along these extensions (Fig. 3b and Supplementary Fig. 8c). Importantly, such particles are an intrinsic property of the SHH molecule, as they are not formed by expression of cytoplasmic, palmitoylated or cholesterol-modified eGFP (data not shown). SHH particles only travel to specific subsets of filopodia emanating from the same cell, revealing tight selectivity and regulation over this process (Fig. 3c).

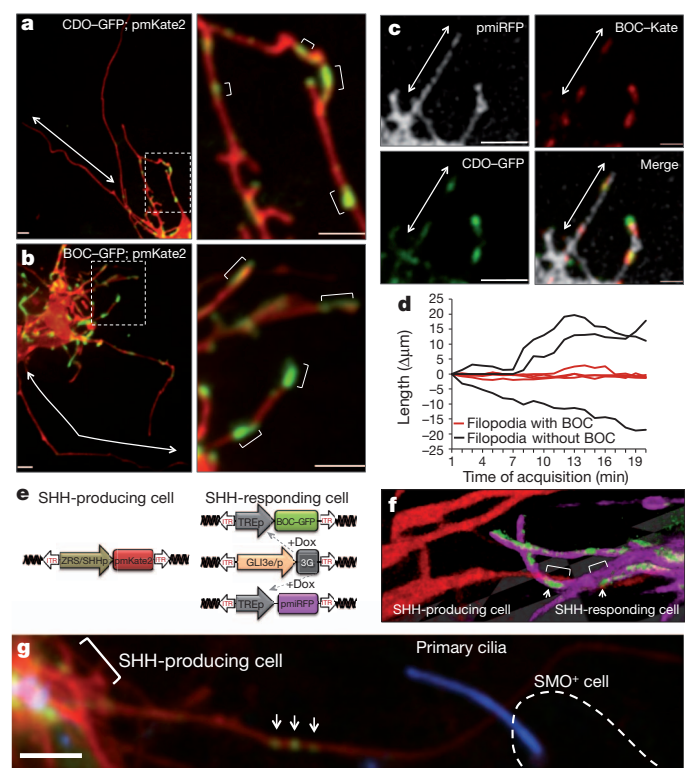
High-speed, real-time imaging showed that SHH<sub>N</sub>-eGFP particles move in both anterograde and retrograde directions along filopodia (Fig. 3c, d and Supplementary Video 8), with a statistically significant net anterograde movement away from the cell body ( $P < 0.002$ ) (Fig. 3e and Supplementary Fig. 9). The maximum velocity of anterograde particle movement,  $120 \text{ nm s}^{-1}$ , is consistent with actin-based myosin motors<sup>17</sup>. Moreover, filopodia containing SHH particles are stabilized and less dynamic than non-particle-containing filopodia (Fig. 3f;  $P < 0.001$ ). These findings reveal that filopodia can distribute SHH ligand at a distance from the cell body. It remains to be determined whether the SHH cholesterol modification, which has been proposed to either promote<sup>18</sup> or restrict<sup>16</sup> the spread of SHH, has additional functions in filopodial transport. To our knowledge, this is the first *in vivo* demonstration of SHH ligand production and movement, revealing an unexpected role for filopodia in this process.

To determine the precise localization of SHH<sub>N</sub>-eGFP within filopodia, we used an optimized split GFP complementation system, consisting of the spGFP<sub>1-10</sub> and spGFP<sub>11</sub> non-fluorescent GFP fragments that reconstitute a fluorescent GFP signal<sup>19</sup>. Notably, SHH<sub>N</sub>-spGFP<sub>11</sub> can physically interact to reconstitute an extracellular leaflet-associated glycosylphosphatidylinositol (GPI)-anchored spGFP<sub>1-10</sub>, producing a GFP signal along the extracellular surface of limb mesenchyme filopodia *in vivo* (Supplementary Fig. 10a, d). We also cultured mesenchymal cells after electroporation of SHH<sub>N</sub>-spGFP<sub>11</sub> and applied a synthesized spGFP<sub>1-10</sub> peptide exogenously to the media (Methods) that similarly produced a GFP fluorescent signal along the filopodial membrane, but not when cells do not express SHH<sub>N</sub>-spGFP<sub>11</sub> or express a control GFP<sub>1-10</sub> cytoplasmic fragment (Supplementary Fig. 10b–d). In addition, ectopic expression of SHH<sub>N</sub>-spGFP<sub>11</sub> tethered to the membrane as a result of its interaction with GPI-spGFP<sub>1-10</sub> leads to the ectopic expansion of PTC1 expression (Supplementary Fig. 10e). Although the split GFP complementation system displays remarkable antibody affinity for GFP fragments<sup>20</sup>, we cannot exclude the possibility that ectopic activation of PTC1 may also derive from a freely diffusible form of SHH<sub>N</sub>-spGFP<sub>11</sub> that we cannot detect. Collectively, these experiments suggest that SHH<sub>N</sub>-eGFP is localized to the extracellular leaflet of the filopodial membrane, where it is competent to signal and interact with SHH receptors.

We next visualized additional SHH signalling components involved in the reception of SHH *in vivo*. In addition to PTC1, which serves as the identified receptor for SHH<sup>21</sup>, additional co-receptors of SHH, including the transmembrane proteins CDO and BOC, are necessary for long-range SHH signalling<sup>22,23</sup>. Interestingly, in the limb bud these co-receptors are only expressed in SHH-responding cells<sup>24</sup>. In contrast to other components of the SHH signalling pathway, such as PTC1–yellow fluorescent protein (YFP) and smoothened (SMO)–GFP that can localize to primary cilia (Supplementary Fig. 11), our live imaging reveals that CDO-GFP and BOC-GFP do not. Instead, they exhibit marked localization to discrete microdomains along long filopodial (average size  $2.4 \pm 1.5 \mu\text{m}$ ; range  $0.6\text{--}8.5 \mu\text{m}$ ) that remain static and display little lateral movement (Fig. 4a, b). Moreover, there is substantial co-localization of these co-receptors to these microdomains in only a subset of filopodia emanating from individual SHH-responding cells, reflecting tight spatial regulation (Fig. 4c). There is a statistically significant stabilization of the filopodia containing SHH co-receptors (Fig. 4d;  $P < 0.001$ ). The molecular roles of CDO and BOC in transducing long-range signalling have been poorly

understood, although their binding to SHH is independent from PTC1 (ref. 25). The unexpected co-localization of these co-receptors on microdomains along filopodia suggests that they may participate in relaying activation of the pathway at a distance from the cell soma.

To mark SHH-producing and -responding cells simultaneously and precisely, we further designed a hybrid GLI3 enhancer and promoter element (Methods and Fig. 4e). Using this dual-expression system, we find that filopodia of SHH-producing cells directly interact with filopodia of SHH-responding cells that contain microdomains of BOC-eGFP (Fig. 4f). Similar stabilized interactions are evident between filopodia containing SHH ligand and SHH-responding cells that have undergone pathway activation as revealed by SMO–blue fluorescent protein (BFP) localization to cilia (Fig. 4g). Together, these results reveal that SHH-producing and -responding cells interact at a distance through filopodial membrane-to-membrane contacts containing SHH ligand and co-receptors.



**Figure 4 | Filopodia on SHH-responding cells display an exquisite distribution and co-localization of SHH co-receptors that interact with SHH-producing filopodia.** **a, b**, Live imaging of CDO-GFP (**a**) and BOC-GFP (**b**) expression in defined microdomains along the filopodial membrane, within subsets of filopodia but not others (arrows). Higher magnification images (**a, b**, right) show multiple positive microdomains of co-receptor localization (brackets) interspersed along the filopodial membrane. **c**, CDO-GFP and BOC-GFP are co-localized along microdomains of filopodia (arrows) labelled with membrane-associated near-infrared fluorescent protein (pmiRFP). **d**, BOC-containing filopodia (red lines) are significantly more stabilized than filopodia without BOC (black lines);  $P < 0.001$ ,  $n = 160$  time points. **e**, Expression system to label SHH-producing (left) and SHH-responding (right) cells specifically in the same limb bud. GLI3e/p denotes GLI3 enhancer/promoter. **f**, Representative three-dimensional image of a filopodium from a SHH-producing cell (pmKate2, red) that interacts with domains of BOC-GFP (green) along the filopodia membrane (pmiRFP, fuchsia) of an SHH-responding cell. Arrows show interaction along BOC microdomains (brackets). **g**, SHH-producing cell, indicated by pmKate2 and marked by bracket, with a long filopodium containing SHH<sub>N</sub>-eGFP particles (arrows) that contacts a smoothened-positive cell (SMO<sup>+</sup>; outlined by a dashed line). Smoothened-BFP localization to the cilium is a marker of SHH pathway activation. All scale bars,  $3 \mu\text{m}$ .



Our live imaging studies have identified a specialized class of filopodia with distinct cytoskeletal features that localize and transport components of the SHH signalling pathway, uncovering an important mechanism for the distribution of signalling molecules within tissues. The highly stabilized interactions between filopodia containing SHH ligand and those containing SHH co-receptors strongly suggest that long-range activation of signalling may be mediated through direct receptor–ligand interactions between cell membranes at a distance. Indeed, SHH-producing and -responding cells can extend filopodia as long as 150 µm in length encompassing the entirety of the 300-µm field of SHH signalling<sup>18</sup> within the limb bud. Interestingly, mesenchymal filopodia may share certain properties with those of cellular extensions previously described within invertebrate embryos<sup>5,26</sup>, and these findings are consistent with earlier studies including electron microscopy carried out more than 30 years ago describing the presence of fine cellular extensions on limb bud mesenchyme<sup>27,28</sup>. Future studies will be required to determine whether such ‘specialized filopodia’ are an inherent feature of many, additional cell types that may have escaped previous detection in fixed and stained samples. Moreover, an outstanding question is whether they rely on unique cellular machinery for their generation, composition and ability to transport signalling molecules. For example, our findings revealing that SHH is produced in the form of a particle, which travels along filopodial extensions in a highly directional manner, suggest that a yet unidentified molecular motor may be responsible for the movement of SHH along these structures. It is intriguing to speculate that the movement of proteins, and perhaps other molecules such as nucleic acids, along specialized filopodial networks offers a new mechanism for controlling the precise delivery of molecular information among cells during vertebrate embryonic development, regeneration and pathological processes such as cancer metastasis. Considering the diverse cellular milieus in which signalling molecules have been shown to act, specialized filopodia may be a more adapted feature of certain signalling centres that operate alone or together with other proposed models for ligand distribution, including free diffusion through extracellular space.

## METHODS SUMMARY

A spatially and temporally regulated piggyBac transposition system was developed to express fluorescent fusion proteins in a heritable fashion within the developing chick embryonic limb bud. The regulated expression of introduced transgenes was stable throughout embryogenesis and did not perturb normal limb development. Cellular morphology and characterization of specialized filopodia was assessed through the use of membrane-associated fluorescent proteins and selected markers of the actin and tubulin cytoskeleton. Molecular components of the SHH signalling pathway were visualized as fusion proteins under endogenous regulatory control. High-resolution live imaging of the embryonic limb bud was performed on custom-designed spinning disk confocal systems, allowing for the visualization of multiple fluorescently coupled proteins in real time, at high spatial resolution.

**Full Methods** and any associated references are available in the online version of the paper.

**Received 8 May 2012; accepted 8 April 2013.**

**Published online 28 April 2013.**

1. Zhu, A. J. & Scott, M. P. Incredible journey: how do developmental signals travel through tissue? *Genes Dev.* **18**, 2985–2997 (2004).
2. Niswander, L. Pattern formation: old models out on a limb. *Nature Rev. Genet.* **4**, 133–143 (2003).
3. Hsiung, F., Ramirez-Weber, F.-A., Iwaki, D. D. & Kornberg, T. B. Dependence of *Drosophila* wing imaginal disc cytonemes on Decapentaplegic. *Nature* **437**, 560–563 (2005).
4. Ramirez-Weber, F. A. & Kornberg, T. B. Cytonemes: cellular processes that project to the principal signaling center in *Drosophila* imaginal discs. *Cell* **97**, 599–607 (1999).

5. Roy, S., Hsiung, F. & Kornberg, T. B. Specificity of *Drosophila* cytonemes for distinct signaling pathways. *Science* **332**, 354–358 (2011).
6. Riddle, R. D., Johnson, R. L., Laufer, E. & Tabin, C. Sonic hedgehog mediates the polarizing activity of the ZPA. *Cell* **75**, 1401–1416 (1993).
7. Yang, Y. *et al.* Relationship between dose, distance and time in Sonic Hedgehog-mediated regulation of anteroposterior polarity in the chick limb. *Development* **124**, 4393–4404 (1997).
8. Yusa, K., Rad, R., Takeda, J. & Bradley, A. Generation of transgene-free induced pluripotent mouse stem cells by the piggyBac transposon. *Nature Methods* **6**, 363–369 (2009).
9. Kerber, M. L. & Cheney, R. E. Myosin-X: a MyTH-FERM myosin at the tips of filopodia. *J. Cell Sci.* **124**, 3733–3741 (2011).
10. Mogilner, A. & Rubinstein, B. The physics of filopodial protrusion. *Biophys. J.* **89**, 782–795 (2005).
11. Munsie, L. N., Caron, N., Desmond, C. R. & Truant, R. Lifeact cannot visualize some forms of stress-induced twisted f-actin. *Nature Methods* **6**, 317 (2009).
12. Breitsprecher, D. *et al.* Cofilin cooperates with fascin to disassemble filopodial actin filaments. *J. Cell Sci.* **124**, 3305–3318 (2011).
13. Niswander, L., Jeffrey, S., Martin, G. R. & Tickle, C. A positive feedback loop coordinates growth and patterning in the vertebrate limb. *Nature* **371**, 609–612 (1994).
14. Maas, S. A., Suzuki, T. & Fallon, J. F. Identification of spontaneous mutations within the long-range limb-specific *Sonic hedgehog* enhancer (ZRS) that alter *Sonic hedgehog* expression in the chicken limb mutants *oligozeugodactyly* and *silkie* breed. *Dev. Dyn.* **240**, 1212–1222 (2011).
15. Ingham, P. W. Hedgehog signaling: a tale of two lipids. *Science* **294**, 1879–1881 (2001).
16. Li, Y., Zhang, H., Litingtung, Y. & Chiang, C. Cholesterol modification restricts the spread of SHH gradient in the limb bud. *Proc. Natl Acad. Sci. USA* **103**, 6548–6553 (2006).
17. Berg, J. S. & Cheney, R. E. Myosin-X is an unconventional myosin that undergoes intrafilopodial motility. *Nature Cell Biol.* **4**, 246–250 (2002).
18. Lewis, P. M. *et al.* Cholesterol modification of sonic hedgehog is required for long-range signaling activity and effective modulation of signaling by Ptc1. *Cell* **105**, 599–612 (2001).
19. Cabantous, S., Terwilliger, T. C. & Waldo, G. S. Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nature Biotechnol.* **23**, 102–107 (2005).
20. Kaddour, L., Magdeleine, E., Waldo, G. S., Joly, E. & Cabantous, S. One-step split GFP staining for sensitive protein detection and localization in mammalian cells. *Biotechniques* **49**, 727–736 (2010).
21. Marigo, V., Davey, R. A., Zuo, Y., Cunningham, J. M. & Tabin, C. J. Biochemical evidence that patched is the Hedgehog receptor. *Nature* **384**, 176–179 (1996).
22. Allen, B. L. *et al.* Overlapping roles and collective requirement for the coreceptors GAS1, CDO, and BOC in SHH pathway function. *Dev. Cell* **20**, 775–787 (2011).
23. Kavran, J. M., Ward, M. D., Oladosu, O. O., Mulepati, S. & Leahy, D. J. All mammalian Hedgehog proteins interact with cell adhesion molecule, down-regulated by oncogenes (CDO) and brother of CDO (BOC) in a conserved manner. *J. Biol. Chem.* **285**, 24584–24590 (2010).
24. Tenzen, T. *et al.* The cell surface membrane proteins Cdo and Boc are components and targets of the Hedgehog signaling pathway and feedback network in mice. *Dev. Cell* **10**, 647–656 (2006).
25. Yao, S., Lum, L. & Beachy, P. The Ihog cell-surface proteins bind Hedgehog and mediate pathway activation. *Cell* **125**, 343–357 (2006).
26. Miller, J., Fraser, S. E. & McClay, D. Dynamics of thin filopodia during sea urchin gastrulation. *Development* **121**, 2501–2511 (1995).
27. Boehm, B. *et al.* The role of spatially controlled cell proliferation in limb bud morphogenesis. *PLoS Biol.* **8**, e1000420 (2010).
28. Kelley, R. O. & Fallon, J. F. Identification and distribution of gap junctions in the mesoderm of the developing chick limb bud. *J. Embryol. Exp. Morphol.* **46**, 99–110 (1978).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank D. Mullins for discussion on the actin cytoskeleton, as well as G. Martin and members of the Barna laboratory for discussion and critical reading of the manuscript. We thank K. Cabaltera for technical assistance. This work was supported by Spanish Ministry of Education and Science (E.L.), Program for Breakthrough Biomedical Research, UCSF (M.B.), the March of Dimes Basil O'Connor Scholar Research Award (M.B.), and the National Institute of Arthritis and Musculoskeletal and Skin Disease, part of NIH, under award number NIH R21AR062262 (M.B.).

**Author Contributions** M.B. conceived and supervised the project; T.A.S., E.L. and M.B. designed experiments; T.A.S. and E.L. performed experiments. All authors analysed the data, critically discussed the results, and contributed towards the writing and preparation of the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.B. ([mbarna@stanford.edu](mailto:mbarna@stanford.edu)).



## METHODS

**Plasmid expression constructs.** For extended expression during chicken embryogenesis, the piggyBac transposition system was used. The parental plasmid pCAG-EBNXN<sup>8</sup> containing the minimal 5' (314 base pairs (bp)) and 3' (242 bp) ITRs of the piggyBac transposon<sup>29</sup> was used as the plasmid backbone for expression experiments. PBX contains the cytomegalovirus (CMV) enhancer chicken  $\beta$ -actin (CAG) promoter expression cassette modified to include the Invitrogen Gateway RfA cassette, allowing for phiC31-mediated recombination from Gateway Entry vectors. For doxycycline-inducible expression, a modified TRE-3G enhancer minimal promoter element (Clontech) replaced the CAG cassette of PBX to generate PB-TREX. For the transposase-mediated insertion of the piggyBac transposon cassette, the improved piggyBac transposase<sup>30</sup> was expressed via the CAG promoter in the plasmid CAGEN (C. Cepko, Addgene plasmid 11160; ref. 31). Fluorescent reporter and signalling constructs were introduced into the Gateway system through high-fidelity amplification and insertion into pENTR/DTOPO (Invitrogen) as an intermediate to facilitate cloning. The integrity of all constructs was confirmed by DNA sequencing.

**Membrane-associated fluorescent proteins.** For multicolour labelling of cells and signalling components, the monomeric fluorescent proteins TagBFP (Evrogen), monomeric eGFP (K. Svoboda, Addgene plasmid 18696; ref. 32), superfolder GFP (sfGFP)<sup>33</sup>, mKate2 (Evrogen) and oligomeric iRFP (V. Verkhusha, Addgene plasmid 31857; ref. 34) were selected based on their spectral characteristics as well as their suitability for live imaging of fusion proteins. Inner leaflet membrane-associated palmitoylated fluorescent proteins were generated by the addition of the 20-amino-acid sequence of rat GAP-43 MLCCMRRTKQVEKNDEQKI to the N terminus of each individual fluorescent protein through sequential PCR amplification; these constructs are designated pm-XFP.

**Mosaic expression of membrane fluorescent proteins.** For Cre-mediated recombination experiments, a loxP-SV40 pA stop-loxP cassette (LSL) (D. Stainier, Addgene 24334; ref. 35) was introduced into the parental ubiquitous and inducible promoter constructs (PB and PB-TRE, respectively). pmKate2 and pmeGFP were placed before and after the LSL cassette, respectively, to generate a reporter construct for Cre activity. To generate a self-inactivating Cre recombinase construct, Cre recombinase containing an N-terminal nuclear localization signal was placed between the lox sites in the PB-LSL vector.

**Subcellular markers.** To label F-actin, two markers were used that show improved cell viability relative to eGFP-actin, UCHD-eGFP (D. Mullins) and LifeAct-mKate2. LifeAct-mKate2 was generated by the addition of the 17-amino-acid sequence of Abp140 from *Saccharomyces cerevisiae* (underlined) and linker sequence MGVDLIKKFESISKEEGDPPVAT to the N terminus of mKate2 (refs 36, 37). Additional markers of the actin cytoskeleton including bovine myosin X-heavy meromyosin-eGFP (R. Cheney<sup>38</sup>), human moesin-eGFP (S. Shaw, Addgene plasmid 20671; ref. 39), human cofilin-eGFP (J. Bamberg) and human fascin-eGFP (D. Vignjevic) were used. Markers of the tubulin cytoskeleton included TAU-GFP (P. Mombaerts) and EB3-Wasabi (Allele Biosciences). Human ARL13B-mKate2 (J. Reiter) was used as a marker of cilia. Cholesterol-modified eGFP was constructed placing amino acids 197–437 of mouse SHH on the carboxy-terminal portion of eGFP. The above fluorescent protein constructs were amplified by high-fidelity PCR, inserted into the pENTR/DTOPO vector and subsequently cloned into the piggyBac expression construct.

**SHH signalling pathway fusion proteins.** Fluorescent fusion proteins were amplified by high-fidelity PCR or overlap-extension PCR, inserted in the pENTR/DTOPO vector and subsequently cloned into a piggyBac expression construct (see above). Mouse *Shh* cDNA was provided by A. McMahon. SHH<sub>p</sub>-eGFP was generated by placing monomeric eGFP between Ser 196 and Gly 197 of mouse SHH through overlap-extension PCR. We optimized the position of monomeric GFP coding sequence relative to the SHH cholesterol modification site to improve *in vivo* expression<sup>40–42</sup>. This resulted in an improved GFP signal *in vivo*; however, SHH<sub>p</sub>-eGFP was less intense than SHH<sub>N</sub>-eGFP and other fusion proteins, which may reflect inefficient processing or stability issues<sup>43</sup>. SHH<sub>N</sub>-eGFP, a fusion protein lacking the C-terminal proteolytic domain and resulting cholesterol addition, was generated by deleting amino acids 197 to 437 after the eGFP fusion. CDO-GFP and BOC-GFP were made as C-terminal fusion proteins<sup>44</sup>. Mouse BOC-mKate2 was generated by replacing the cytoplasmic tail GFP with mKate2. Murine SMO-TagBFP was generated replacing the fluorophore of SMO-GFP (J. Reiter). PTC1-YFP was from J. Reiter.

**spGFP complementation system for SHH.** The optimized split GFP complementation system described previously<sup>19</sup> was used to assess whether SHH<sub>N</sub>-eGFP was localized to the extracellular surface of filopodia. SHH<sub>N</sub>-spGFP<sub>11</sub> was generated by placing the M3 peptide after amino acid 196 of mouse SHH separated by a flexible linker, (GGGS)×3. SHH<sub>N</sub>-spGFP<sub>11</sub> and GPI-linked CD14-spGFP<sub>1–10</sub> (ref. 45) were inserted into the PB-TRE expression construct. There was no

detectable fluorescence observed with either construct independently consistent with previous studies<sup>19</sup>.

To address further the subcellular localization of SHH in relation to mesenchymal filopodia, we used the exogenous application of a commercially produced spGFP<sub>1–10</sub> (Sandia Laboratories)<sup>20</sup>. In brief, after the electroporation of the chick embryo somatic lateral plate mesoderm, *ex vivo* cultures of limb mesenchymal cells were prepared as described previously<sup>46</sup>. After plating of these cells, the purified spGFP<sub>1–10</sub> fragment was added exogenously to the live culture and imaged. Positive interaction between the SHH<sub>N</sub>-spGFP<sub>11</sub> and the spGFP<sub>1–10</sub> fragments was observed as a fluorescent signal 4 h after reagent addition. As a control for possible endocytosis of the spGFP<sub>1–10</sub> fragment, a spGFP<sub>11</sub> luciferase fusion protein localized to the cytoplasm did not produce a fluorescent signal.

***G. gallus* SHH promoter and enhancer element.** The 1.7-kilobase (kb) SHH limb specific enhancer element designated the ZRS, and the 1.1-kb chicken SHH minimal enhancer element<sup>44</sup> were cloned into the PBX vector replacing the CAG promoter cassette to generate a Gateway compatible expression construct. Subsequently, pmEGFP or the tetracycline activator protein 3G (Clontech) were inserted into the expression cassette to allow for either constitutive or doxycycline-inducible spatially restricted expression in the ZPA.

**GLI3 intronic enhancer and promoter element.** To allow for fluorescent transgenes to be expressed in SHH-responding cells of the limb bud, a screen for promoter elements that would correctly regulate expression was performed. In brief, several human enhancer elements that were previously identified following chromatin immunoprecipitation DNA sequencing (ChIP-seq) of the enhancer associated p300 protein and mouse transgenic analysis<sup>47</sup> were subsequently tested in chick embryos when coupled to various minimal promoter elements, HSP68, E1b, thymidine kinase and minimal CMV. Specific expression was achieved when the human GLI3 intronic enhancer element, hs1586 (ref. 48), was coupled with the rat minimal Gli3 promoter<sup>49</sup>. Subsequently, pmKate2 (Supplementary Fig. 12) or 3G (Clontech) were inserted into the expression cassette to allow for either constitutive or doxycycline-inducible spatially restricted expression in the GLI3-expressing and SHH-responding cells of the limb.

**Chicken embryo manipulation and electroporation.** Fertilized chicken eggs (*G. gallus*) were purchased from Petaluma Farms and subsequently stored at 16 °C. Eggs were incubated in a non-rotary incubator at 38.5 °C until the desired stage according to Hamburger and Hamilton (HH)<sup>50</sup>. Stage HH13–15 chick embryos were windowed following standard techniques in preparation for electroporation<sup>51</sup>. Embryos were visualized with the assistance of a 470/40 nm band-width emission filter, which provided the necessary contrast for injection. PBS without Ca<sup>2+</sup>/Mg<sup>2+</sup> was applied to the embryo. The vitelline membrane above the forelimb field was carefully sub-dissected, and additional solution was placed over the embryo. DNA constructs, with a final concentration 1–5  $\mu$ g  $\mu$ l<sup>-1</sup>, diluted in endotoxin-free H<sub>2</sub>O were combined with phenol red (0.1 mM final concentration) to aid in visualization. A 1.0-mm inner diameter capillary glass electrode was backfilled with DNA injection solution and a volume of solution was pressure injected (WPI Picopump) into the embryonic coelom, to fill completely the anterior to posterior extent of the forelimb territory. For the negative electrode, a 250- $\mu$ m diameter platinum rod with a 4-mm length and 2-mm exposed surface (Nepagene) was inserted into the yolk and positioned beneath the forelimb field, approximately 0.5–1 mm below the embryo. A 250- $\mu$ m diameter platinum rod with a 1-mm exposed tip served as the positive electrode and was positioned above the forelimb field with an approximate distance of 2 mm. A square-wave pulse train consisting of 8 V, three pulses, 50-ms duration with a 1-s interpulse interval was delivered via a Nepa 21 electroporator (Nepagene). This delivery resulted in an approximate current of 8–14 mA with energy of 10–18 mJ. Embryos were returned to 37.5 °C for the remainder of the incubation period. In experiments using the piggyBac transposition system, a 1:5–1:10 molar ratio of piggyBac transposase helper plasmid, HypBase, was combined with the transposon expression construct to mediate integration and high-level expression. This ratio resulted in persistent expression in the embryonic limb through embryonic day (E)12, 10 days after electroporation (Supplementary Fig. 1a and unpublished observations). Moreover, limb development and resulting morphology was normal as assessed with Alcian Blue cartilage staining (Supplementary Fig. 1b). For induction of gene expression with the inducible 3G system, 12–24 h before imaging, 50 ng doxycycline (Clontech) in 500  $\mu$ l in HBSS was injected beneath the embryonic vasculature. For experiments with iRFP as a fluorescent protein, 75 ng biliverdin (Frontier Scientific) was administered more than 4 h before imaging.

**Live imaging of embryos.** To facilitate accessibility of the chick embryo to live imaging with minimal perturbation, embryos were cultured *ex ovo* with improved viability and sustainability<sup>52</sup>. In brief, eggs after 48 h of incubation at 38.5 °C were prepared in sterile fashion and the embryo was directly transferred to a sterile autoclaved 60-mm diameter 35-mm depth crystallization dish by cracking the egg and allowing the albumin and yolk to fall gently into the vessel. Five millilitres of

sterile PBS without  $\text{Ca}^{2+}/\text{Mg}^{2+}$  containing  $10\times$  penicillin/streptomycin solution was added to prevent dehydration. The embryo was subsequently covered with a vented sterile lid and placed at  $37.5^\circ\text{C}$ . Electroporation was performed as described above. For live imaging on a confocal Axio examiner system (see below), a custom-heated stage top incubator (BioOptechs) was designed allowing for the insertion of a water dipping objective for continuous imaging while maintaining temperature, humidity and normal growth of chick embryos (Supplementary Fig. 1c).

For imaging with the Zeiss Axio Observer confocal system, mouse embryos or electroporated chick embryos were collected into imaging media (DMEM/F12 with HEPES without phenol red containing 10% heat-inactivated FBS; Invitrogen). Extraembryonic membranes were carefully removed and the entire embryo or the isolated forelimb was transferred and positioned on a 35-mm glass bottom culture dish containing a 14-mm German glass coverslip as its base (MatTek Corporation). A 12-mm coverslip was placed above the embryo secured on a ring of Vaseline that served to elevate the coverslip from the embryo, this placement was done to limit the movement of the limb bud during imaging. The chamber containing the embryo or limb bud was placed in a  $37^\circ\text{C}$  heated microscope incubator (Solent Scientific) and imaged as described below.

**Mice.** The *mT/mG* (ref. 53), *Shh<sup>CreERT2/+</sup>* (ref. 54) and *Gli<sup>CreERT2/+</sup>* (ref. 55) mice were purchased from Jackson laboratories and maintained on C57BL/6J background. The *mT/mG* transgenic line is a double-fluorescent Cre reporter mouse that expresses tandem dimer Tomato (*mT*) before Cre-mediated excision and membrane-targeted enhanced green fluorescent protein (*mG*) after excision to mark defined populations of limb mesenchymal cells. For induction with tamoxifen, 4 mg of tamoxifen dissolved in corn oil (Sigma) was delivered via orogastric gavage at E9.5 and E10.5 for *Shh<sup>CreERT2/+</sup>*; *mT/mG<sup>+/+</sup>* and *Gli<sup>CreERT2/+</sup>*; *mT/mG<sup>+/+</sup>* mice, respectively. For deletion of *Cdc42* experiments<sup>56</sup>, in which *Cdc42* is critically required for filopodia formation *in vitro*<sup>57</sup>, tamoxifen was delivered on E9.5 for *Gli<sup>CreERT2/+</sup>*; *mT/mG<sup>+/+</sup>*; *Cdc42<sup>loxP/loxP</sup>* mice. All animals were maintained at The University of California, San Francisco, and procedures were performed using Institutional Animal Care and Use Committees (IACUC)-approved protocols that adhere to the standards of the NIH.

**Confocal microscopy.** Images were primarily acquired on one of two custom-built spinning disk microscopes. The Zeiss Axio Observer Microscopy system is coupled to a Perkin Elmer UltraVIEW Vox spinning disk confocal microscopy system. The UltraVIEW Vox system used 405 nm, 488 nm, 561 nm and 640 nm solid state laser lines paired with emission filters (Semrock) that were specifically selected to minimize crosstalk across various fluorescent proteins when excited with appropriate solid state sources. Images were captured in the gain mode of a back-thinned electron multiplying CCD camera (Hamamatsu ImageEM C91003). Acquisition of images was accomplished with the Velocity Acquisition suite for multidimensional multichannel time-lapse recordings. Laser power and exposure settings were adjusted to minimize phototoxicity during the sequence acquisition, power output measured at the entrance to the spinning disk of less than 2.5 mW for all excitation wavelengths.

*In ovo* imaging was performed on a Zeiss Axio Examiner Microscopy System with a continuous zoom system coupled to a custom built confocal system, with a modified Yokogawa CSU10 scan head that contains an additional patterned array of microlenses. A Hamamatsu ImageEM gain CCD camera was used to capture multichannel and multidimensional images. A custom-built heated stage top incubator allowed for precise insertion of the water-dipping objective for continuous imaging of chick embryos in a micro-controlled environment (Supplementary Fig. 1, see live imaging of embryos).

**Image analysis.** Acquired images were processed through the use of the Velocity 6.0 Visualization and Quantification suite (Perkin Elmer). Image size calibrations were performed for each objective and optovar setting with a calibrated graticule (Electron Microscopy Sciences). Manual measurements of filopodia, fluorescent puncta and trajectories were obtained across the *z* space of a three-dimensional stack as well as in time for continuous imaging. Velocity-recorded values were transferred and processed in Microsoft Excel for subsequent quantification and presentation. For determining filopodia orientation along limb axes, the relative bearing of each filopodial extension across mouse genotypes was categorized into four quadrants corresponding to the anterior, distal, posterior and proximal axes. This quadrant categorization, encompassing  $45^\circ$  around the centre point, allowed for the comparison of vector orientations. Statistical analysis was performed assuming homogeneous distribution of variances and applying Student's *t*-test between percentages of anterior and posterior orientation versus proximal and distal orientation per cell. For the measurement of fluorescent intensities values, maximal cellular intensity values were obtained with the subsequent subtraction of

background fluorescent intensity. Filopodia were manually selected and fluorescent intensity values were determined using Velocity 6.0 Quantification suite. These values were normalized to maximal intensity for the cell to account for differences in protein expression levels. Image presentations were generated in the Velocity 6.0 Visualization suite. For select multicolour confocal acquisitions, deconvolution using the calculated point spread function was applied with the Velocity 6.0 Restoration package.

29. Li, X. *et al.* piggyBac internal sequences are necessary for efficient transformation of target genomes. *Insect Mol. Biol.* **14**, 17–30 (2005).
30. Yusa, K., Zhou, L., Li, M. A., Bradley, A. & Craig, N. L. A hyperactive piggyBac transposase for mammalian applications. *Proc. Natl Acad. Sci. USA* **108**, 1531–1536 (2011).
31. Matsuda, T. & Cepko, C. L. Electroporation and RNA interference in the rodent retina *in vivo* and *in vitro*. *Proc. Natl Acad. Sci. USA* **101**, 16–22 (2004).
32. Harvey, C. D., Yasuda, R., Zhong, H. & Svoboda, K. The spread of Ras activity triggered by activation of a single dendritic spine. *Science* **321**, 136–140 (2008).
33. Pédélecq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nature Biotechnol.* **24**, 79–88 (2005).
34. Filonov, G. S. *et al.* Bright and stable near-infrared fluorescent protein for *in vivo* imaging. *Nature Biotechnol.* **29**, 757–761 (2011).
35. Hesselton, D., Anderson, R. M., Beinart, M. & Stainier, D. Y. R. Distinct populations of quiescent and proliferative pancreatic beta-cells identified by HOTTre mediated labeling. *Proc. Natl Acad. Sci. USA* **106**, 14896–14901 (2009).
36. Riedl, J. *et al.* Lifeact mice for studying F-actin dynamics. *Nature Methods* **7**, 168–169 (2010).
37. Riedl, J. *et al.* Lifeact: a versatile marker to visualize F-actin. *Nature Methods* **5**, 605–607 (2008).
38. Bohil, A. B., Robertson, B. W. & Cheney, R. E. Myosin-X is a molecular motor that functions in filopodia formation. *Proc. Natl Acad. Sci. USA* **103**, 12411–12416 (2006).
39. Hao, J.-J. *et al.* Phospholipase C-mediated hydrolysis of PIP2 releases ERM proteins from lymphocyte membrane. *J. Cell Biol.* **184**, 451–462 (2009).
40. Callejo, A., Quijada, L. & Guerrero, I. Detecting tagged Hedgehog with intracellular and extracellular immunocytochemistry for functional analysis. *Methods Mol. Biol.* **397**, 91–103 (2007).
41. Vincent, S., Thomas, A., Brasher, B. & Benson, J. D. Targeting of proteins to membranes through hedgehog auto-processing. *Nature Biotechnol.* **21**, 936–940 (2003).
42. Vyas, N. *et al.* Nanoscale organization of hedgehog is essential for long-range signaling. *Cell* **133**, 1214–1227 (2008).
43. Chamberlain, C. E., Jeong, J., Guo, C., Allen, B. L. & McMahon, A. P. Notochord-derived SHH concentrates in close association with the apically positioned basal body in neural target cells and forms a dynamic gradient during neural patterning. *Development* **135**, 1097–1106 (2008).
44. Okada, A. *et al.* Boc is a receptor for sonic hedgehog in the guidance of commissural axons. *Nature* **444**, 369–373 (2006).
45. Pinaud, F. & Dahan, M. Targeting and imaging single biomolecules in living cells by complementation-activated light microscopy with split-fluorescent proteins. *Proc. Natl Acad. Sci. USA* **108**, E201–E210 (2011).
46. Barna, M. & Niswander, L. Visualization of cartilage formation: insight into cellular properties of skeletal progenitors and chondrodysplasia syndromes. *Dev. Cell* **12**, 931–941 (2007).
47. Visel, A. *et al.* ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**, 854–858 (2009).
48. Visel, A., Minovitsky, S., Dubchak, I. & Pennacchio, L. A. VISTA Enhancer Browser—a database of tissue-specific human enhancers. *Nucleic Acids Res.* **35**, D88–D92 (2007).
49. Cao, D. *et al.* The expression of *Gli3*, regulated by *HOXD13*, may play a role in idiopathic congenital talipes equinovarus. *BMC Musculoskelet. Disord.* **10**, 142 (2009).
50. Hamburger, V. A series of normal stages in the development of the chick embryo. *J. Morphol.* **195**, 231–272 (1951).
51. Krull, C. E. A primer on using *in ovo* electroporation to analyze gene function. *Dev. Dyn.* **229**, 433–439 (2004).
52. Auerbach, R., Kubai, L., Knighton, D. & Folkman, J. A simple procedure for the long-term cultivation of chicken embryos. *Dev. Biol.* **41**, 391–394 (1974).
53. Muzumdar, M. D., Tasic, B., Miyamichi, K., Li, L. & Luo, L. A global double-fluorescent Cre reporter mouse. *Genesis* **45**, 593–605 (2007).
54. Harfe, B. D. *et al.* Evidence for an expansion-based temporal SHH gradient in specifying vertebrate digit identities. *Cell* **118**, 517–528 (2004).
55. Ahn, S. & Joyner, A. L. Dynamic changes in the response of cells to positive hedgehog signaling during mouse limb patterning. *Cell* **118**, 505–516 (2004).
56. Chen, L. *et al.* *Cdc42* deficiency causes Sonic hedgehog-independent holoprosencephaly. *Proc. Natl Acad. Sci. USA* **103**, 16520–16525 (2006).
57. Nobes, C. D. & Hall, A. Rho, rac, and cdc42 GTPases regulate the assembly of multimolecular focal complexes associated with actin stress fibers, lamellipodia, and filopodia. *Cell* **81**, 53–62 (1995).



# Macropinocytosis of protein is an amino acid supply route in Ras-transformed cells

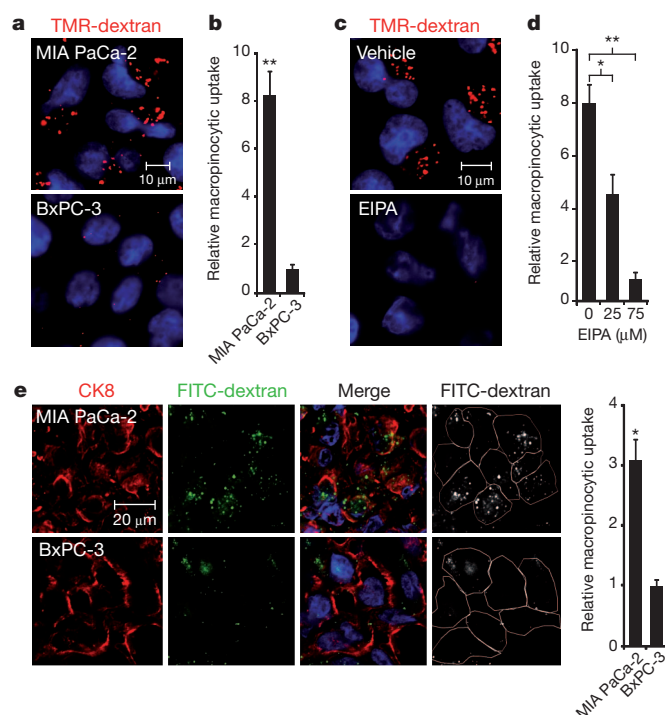
Cosimo Commisso<sup>1</sup>, Shawn M. Davidson<sup>2\*</sup>, Rengin G. Soydaner-Azeloglu<sup>1\*</sup>, Seth J. Parker<sup>3\*</sup>, Jurre J. Kamphorst<sup>4</sup>, Sean Hackett<sup>4</sup>, Elda Grabocka<sup>1</sup>, Michel Nofal<sup>4</sup>, Jeffrey A. Drebin<sup>5</sup>, Craig B. Thompson<sup>6</sup>, Joshua D. Rabinowitz<sup>4</sup>, Christian M. Metallo<sup>3</sup>, Matthew G. Vander Heiden<sup>2,7</sup> & Dafna Bar-Sagi<sup>1</sup>

Macropinocytosis is a highly conserved endocytic process by which extracellular fluid and its contents are internalized into cells through large, heterogeneous vesicles known as macropinosomes. Oncogenic Ras proteins have been shown to stimulate macropinocytosis but the functional contribution of this uptake mechanism to the transformed phenotype remains unknown<sup>1–3</sup>. Here we show that Ras-transformed cells use macropinocytosis to transport extracellular protein into the cell. The internalized protein undergoes proteolytic degradation, yielding amino acids including glutamine that can enter central carbon metabolism. Accordingly, the dependence of Ras-transformed cells on free extracellular glutamine for growth can be suppressed by the macropinocytic uptake of protein. Consistent with macropinocytosis representing an important route of nutrient uptake in tumours, its pharmacological inhibition compromises the growth of Ras-transformed pancreatic tumour xenografts. These results identify macropinocytosis as a mechanism by which cancer cells support their unique metabolic needs and point to the possible exploitation of this process in the design of anticancer therapies.

To date, the induction of macropinocytosis by oncogenic Ras has been characterized in the setting of overexpressed proteins<sup>1–3</sup>. To determine whether stimulated macropinocytosis is a feature of cancer cells endogenously expressing oncogenic Ras, we analysed fluid-phase uptake in human pancreatic and urinary bladder cancer cell lines harbouring oncogenic Ras mutations and compared uptake to wild-type Ras-expressing cells originating from carcinomas of the same tissue type. Macropinosomes were visualized on the basis of the ability of cells to internalize extracellular medium containing tetramethylrhodamine-labelled high-molecular-mass dextran (TMR-dextran), an established marker of macropinocytosis. Pancreatic adenocarcinoma-derived human MIA PaCa-2 cells, which are homozygous for the *KRAS*<sup>G12C</sup> allele<sup>4</sup>, displayed appreciably higher levels of TMR-dextran uptake compared to BxPC-3 cells, which express wild-type *KRAS*<sup>5</sup> (Fig. 1a, b). That the TMR-dextran labelling in the oncogenic Ras-expressing cells reflects uptake through macropinocytosis is indicated by the observation that uptake was inhibited in a dose-dependent manner by 5-(N-ethyl-N-isopropyl)amiloride (EIPA) (Fig. 1c, d), which has been shown to inhibit macropinosome formation without affecting other endocytic pathways<sup>6,7</sup>. Importantly, the knockdown of *KRAS* led to an attenuation of macropinocytosis, confirming the dependence of this uptake mechanism on oncogenic Ras expression (Supplementary Fig. 1a–d). This conclusion is further supported by the observation that bladder carcinoma-derived T24 cells, which are homozygous for the *HRAS*<sup>G12V</sup> allele<sup>8</sup>, exhibit increased levels of macropinocytosis relative to 5637 cells, which express wild-type *HRAS* (Supplementary Fig. 2a–c)<sup>9</sup>.

To examine whether oncogenic Ras-expressing cells engage in macropinocytosis *in vivo*, we used both a heterotopic xenograft mouse

model and an autochthonous mouse model. For the heterotopic model, tumours were injected with fluorescein isothiocyanate-conjugated dextran (FITC-dextran) and intracellular uptake was



**Figure 1 | Oncogenic KRAS-expressing pancreatic cancer cells display increased levels of macropinocytosis both in culture and *in vivo*.** **a**, A macropinocytosis uptake assay using TMR-dextran as a marker of macropinosomes (red) indicates that MIA PaCa-2 cells display increased levels of macropinocytosis compared to BxPC-3 cells. DAPI (4',6-diamidino-2-phenylindole) staining (blue) identifies nuclei. **b**, Quantification of macropinocytic uptake in pancreatic cancer cells. Data are expressed as arbitrary units and are presented relative to the values obtained for BxPC-3 cells. **c**, Macropinocytic uptake in MIA PaCa-2 cells treated with either vehicle (dimethylsulphoxide (DMSO)) or 75 μM EIPA. **d**, Quantification of macropinocytic uptake in MIA PaCa-2 cells treated with 0, 25 or 75 μM EIPA. Data are presented relative to the values obtained for the 75 μM condition. **e**, Visualization and quantification of macropinocytosis *in vivo*. Representative images from sections of FITC-dextran (green)-injected tumour xenografts stained with anti-CK8 (red). Cell boundaries (white outline) were delineated on the basis of CK8 staining. Data are presented relative to the values obtained for the BxPC-3 tumours. For all graphs, error bars indicate mean  $\pm$  s.e.m. for  $n = 3$  independent experiments with at least 300 cells scored per experiment. Statistical significance was determined by Student's *t*-test; \* $P < 0.05$ , \*\* $P < 0.01$ .

<sup>1</sup>Department of Biochemistry and Molecular Pharmacology, New York University School of Medicine, New York, New York 10016, USA. <sup>2</sup>Koch Institute for Integrative Cancer Research and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. <sup>3</sup>Department of Bioengineering, University of California, San Diego, La Jolla, California 92093, USA. <sup>4</sup>Lewis-Sigler Institute for Integrative Genomics, Carl Icahn Laboratory, Princeton University, Princeton, New Jersey 08544, USA. <sup>5</sup>Department of Surgery, Hospital of the University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. <sup>6</sup>Cancer Biology and Genetics Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. <sup>7</sup>Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA.

\*These authors contributed equally to this work.



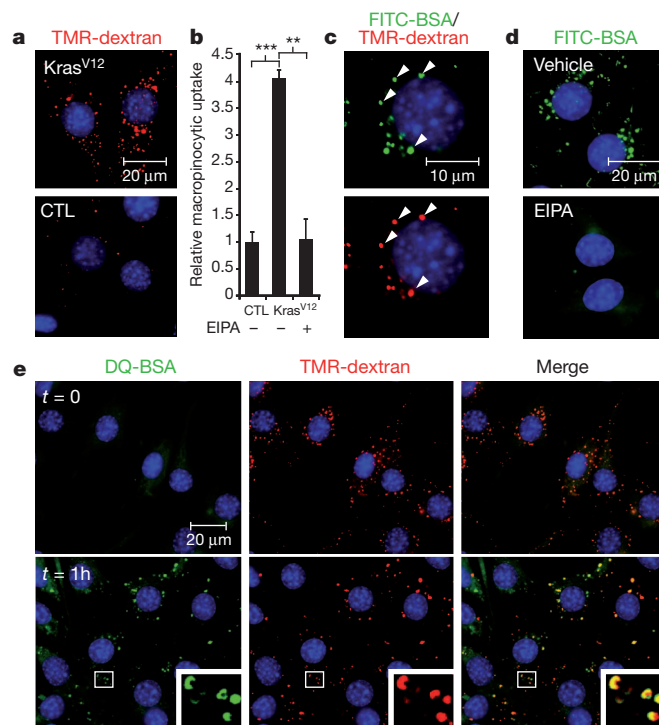
assessed by fluorescent microscopy of tissue sections. The number of macropinosomes identified as FITC-positive puncta was markedly higher in tumours derived from MIA PaCa-2 cells relative to BxPC-3-derived tumours (Fig. 1e). To confirm that the macropinosomes are a feature of the transplanted cells rather than the host cells, the tumour sections were stained with anti-CK8 antibody, which selectively labels the transplanted human epithelial cells<sup>10</sup> (Fig. 1e). To analyse macropinocytosis in mouse pancreatic tumours, we used an autochthonous mouse model of pancreatic cancer<sup>11</sup>. In this model, animals of the genotype *P48-cre; lsl-Kras<sup>G12D</sup>; Trp53<sup>-/+</sup>* (KPC) develop pancreatic intraepithelial neoplasia lesions within 4 weeks of birth and progress to invasive pancreatic ductal adenocarcinoma between 9 and 13 weeks<sup>12</sup>. To assess macropinocytic uptake, KPC mice were injected with FITC-dextran at 12 weeks of age and pancreata were subsequently collected. FITC-positive macropinosomes were detected in CK19-labelled acinoductal cells within mid- to late-stage pancreatic intraepithelial neoplasia lesions of KPC pancreata but not in pancreata from wild-type mice (Supplementary Fig. 3). Altogether, these data indicate that an increased level of macropinocytosis is an attribute of cancer cells expressing oncogenic Ras both *in vitro* and *in vivo*.

In mammals, approximately 70% of the soluble substances found in extracellular fluid can be accounted for by proteins, with serum albumin being the most abundant<sup>13</sup>. Therefore, we sought to determine whether protein uptake through macropinocytosis can be used by oncogenic Ras-expressing cells to meet their metabolic needs for proliferation. As an experimental system, we used oncogenic *Kras*-transformed mouse NIH 3T3 cells because of their documented dependence on glutamine<sup>14</sup>. The expression of oncogenic *Kras<sup>V12</sup>* in these cells was sufficient to stimulate a robust EIPA-sensitive macropinocytic response as measured by TMR-dextran uptake (Fig. 2a, b). To monitor the internalization of albumin, NIH 3T3 *Kras<sup>V12</sup>* cells were incubated with a FITC-labelled form of bovine serum albumin (FITC-BSA). As shown in Fig. 2c, FITC-BSA was incorporated into discrete intracellular structures that co-localized with TMR-dextran. Similar uptake was also observed in MIA PaCa-2 and T24 cells (Supplementary Fig. 4), indicating that albumin internalization can occur through macropinocytosis in cancer cells harbouring endogenous oncogenic Ras mutations. Inhibition of FITC-BSA uptake by treatment with EIPA confirmed that the uptake mechanism was macropinocytosis (Fig. 2d).

The macropinocytic internalization and subsequent degradation of albumin could in principle lead to the generation of amino acids that sustain tumour cell bioenergetics and macromolecular synthesis<sup>15</sup>. To determine whether the internalized albumin is intracellularly degraded, we used a highly self-quenched BODIPY-dye-conjugated form of BSA (DQ-BSA) that emits a bright fluorescent signal only after proteolytic digestion<sup>16</sup>. Dual labelling of cells with DQ-BSA and TMR-dextran was used to establish the macropinocytic origin of the degradative compartment. In NIH 3T3 *Kras<sup>V12</sup>* cells that were immediately fixed following a 30-min incubation with DQ-BSA and TMR-dextran ( $t = 0$ ), there was no appreciable DQ-BSA fluorescence detected in macropinosomes (Fig. 2e). However, in cells that were incubated for 30 min and subsequently chased for 1 h in media free of both DQ-BSA and TMR-dextran, DQ-BSA fluorescence was detected in TMR-positive macropinosomes (Fig. 2e). DQ-BSA fluorescence was also detected within macropinosomes after a 1-h chase in MIA PaCa-2 and T24 cells, indicating that these trafficking events were also occurring in cancer cells harbouring endogenous oncogenic Ras mutations (Supplementary Fig. 5). The degradation of macropinocytosed albumin was dependent on lysosomal hydrolases, as treatment of MIA PaCa-2 and NIH 3T3 *Kras<sup>V12</sup>* cells with bafilomycin A1 prevented the degradation of DQ-BSA (Supplementary Fig. 6). These data demonstrate that oncogenic Ras-expressing cells can harness macropinocytosis for the internalization and degradation of extracellular albumin, and raise the possibility that plasma protein degradation may provide an important source of intracellular amino acids.

Glutamine is a key nutrient for many proliferating cells and is metabolized to glutamate and then  $\alpha$ -ketoglutarate to enter central carbon metabolism<sup>17</sup>. Ras-transformed cells exhibit an increased dependency on glutamine for growth and survival<sup>14,18</sup>. To test whether the degradation of macropinocytosed albumin results in the production of intracellular glutamine, we directly measured the intracellular concentrations of glutamate and  $\alpha$ -ketoglutarate in cells grown either in the absence or presence of albumin. NIH 3T3 *Kras<sup>V12</sup>* cells were cultured for 24 h in complete media supplemented with physiological concentrations of albumin (2 g per 100 ml, 2%). As a control, NIH 3T3 *Kras<sup>V12</sup>* cells were grown in complete media alone. The addition of albumin to the media led to EIPA-sensitive increases in intracellular concentrations of both glutamate and  $\alpha$ -ketoglutarate (Supplementary Fig. 7), indicating that macropinocytic uptake of albumin can increase the levels of glutamate and  $\alpha$ -ketoglutarate in oncogenic Ras-transformed cells.

NIH 3T3 *Kras<sup>V12</sup>* cells use glutamine as a main carbon source for tricarboxylic acid (TCA) cycle anaplerosis<sup>19</sup>. Therefore, to directly trace the fate of protein-derived amino acids in oncogenic Ras-transformed cells, we cultured NIH 3T3 *Kras<sup>V12</sup>* cells in the presence of soluble, heat-inactivated, <sup>13</sup>C-labelled yeast protein. Uptake of



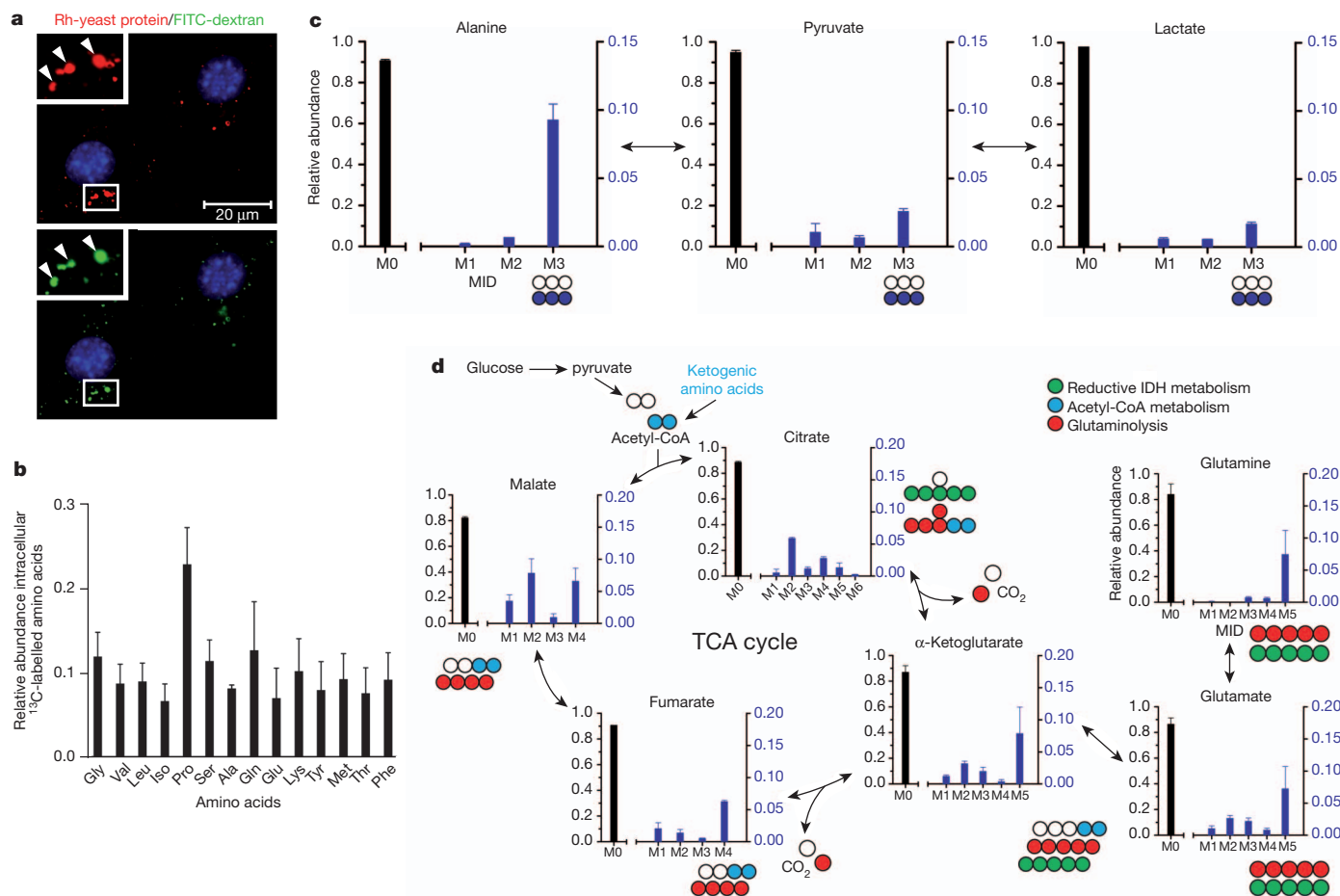
**Figure 2 | Oncogenic Kras-induced macropinocytosis in mouse NIH 3T3 cells mediates the internalization of extracellular albumin, which is subsequently targeted for proteolytic degradation.** **a**, TMR-dextran (red) is internalized at higher levels in NIH 3T3 *Kras<sup>V12</sup>* cells (*Kras<sup>V12</sup>*) compared to untransformed control cells (CTL). **b**, Quantification of macropinocytic uptake in control cells and NIH 3T3 *Kras<sup>V12</sup>* cells incubated with vehicle (DMSO) or with 75  $\mu$ M EIPA. Data are presented relative to the values obtained for the untransformed control cells. Error bars indicate mean  $\pm$  s.e.m. for  $n = 3$  independent experiments with at least 300 cells scored per experiment. Statistical significance was determined by Student's *t*-test; \*\* $P < 0.01$ , \*\*\* $P < 0.001$ . **c**, FITC-BSA (green) is internalized into discrete puncta that co-localize (white arrowheads) with TMR-dextran (red). **d**, FITC-BSA uptake is abrogated by treatment with 75  $\mu$ M EIPA. **e**, Analysis of DQ-BSA fluorescence in NIH 3T3 *Kras<sup>V12</sup>* cells that were co-incubated with DQ-BSA (green) and TMR-dextran (red) and fixed either immediately ( $t = 0$ ) or following a 1-h chase. The fluorescent signal emanating from DQ-BSA ( $t = 1$  h) is an indication of albumin degradation. Insets represent a higher magnification of the boxed areas. Images shown in **c–e** are representative of at least three independent experiments.

yeast-derived protein through macropinocytosis was confirmed microscopically by the fluorescent co-localization of rhodamine-labelled protein and FITC-dextran (Fig. 3a). After 24 h of culture in low-glutamine (0.2 mM) complete medium supplemented with  $^{13}\text{C}$ -labelled yeast proteins, intracellular metabolites were extracted and  $^{13}\text{C}$  labelling was quantified using gas chromatography/mass spectrometry (Supplementary Table 1)<sup>20</sup>. Substantial labelling of numerous intracellular amino acids, including glutamine, was detected (Fig. 3b), whereas only low levels of  $^{13}\text{C}$  amino acids were detected in media samples incubated for the same period of time without cells (Supplementary Fig. 8). Catabolized yeast protein also entered central metabolism, as evidenced by  $^{13}\text{C}$ -labelling of pyruvate, lactate and various TCA metabolites (Fig. 3c, d). Partially labelled mass isotopomers in numerous metabolites suggested that protein-derived amino acids were potentially being metabolized through several pathways, including glutamine anaplerosis/oxidation, acetyl-coenzyme A metabolism, reductive carboxylation and serine/glycine cycling (Fig. 3d and Supplementary Fig. 9).

Increased sensitivity to glutamine deprivation is a hallmark of cancer cells that express oncogenic mutants of Ras<sup>14,18</sup>. Consistent with this phenotype, we found that MIA PaCa-2, T24 and NIH 3T3 Kras<sup>V12</sup>

cells exhibited decreased proliferation at sub-physiological levels of glutamine as measured using the MTT viability assay or by Syto 60 staining (Supplementary Figs 10 and 11). We next examined whether cell growth impairment due to glutamine deprivation could be reversed by culturing cells in the presence of physiological levels of albumin. Albumin supplementation of media containing sub-physiological glutamine concentrations enhanced proliferation, and this effect was abrogated by EIPA treatment (Fig. 4a and Supplementary Fig. 11). The anti-proliferative response to EIPA observed in these settings was rescued by the addition of either extracellular glutamine or  $\alpha$ -ketoglutarate (Fig. 4b). The ability of extracellular albumin to suppress the detrimental effects of glutamine starvation in MIA PaCa-2 cells was dependent on KRAS expression, as KRAS knockdown in these cells diminished the rescuing capacity of albumin (Supplementary Fig. 12). Together, these data indicate that the macropinocytic uptake of albumin could serve to sustain proliferation of oncogenic Ras-transformed cells by constituting a source of glutamine, and potentially other amino acids.

To investigate whether the uptake and degradation of albumin is a unique feature of oncogenic Ras-induced macropinocytosis, we used NIH 3T3 cells expressing Src<sup>Y527F</sup> (NIH 3T3 Src<sup>Y527F</sup>), a constitutively active form of Src<sup>21</sup>. Consistent with previous reports<sup>22–24</sup>,



**Figure 3 | Macropinocytic uptake of extracellular protein drives the accumulation of catabolic intermediates and entry of protein-derived amino acids into central carbon metabolism.** **a**, Rhodamine (Rh)-labelled yeast protein (red) is internalized into puncta (arrowheads) that co-localize with FITC-dextran (green). Insets represent a higher magnification of the boxed areas. **b**, Uniformly  $^{13}\text{C}$ -labelled intracellular amino acid pools were detected in NIH 3T3 Kras<sup>V12</sup> cells after culture in low-glutamine-containing medium (0.2 mM) supplemented with 2%  $^{13}\text{C}$ -labelled yeast protein. **c**, Protein-derived alanine enters central carbon metabolism upon transamination to pyruvate, and pyruvate can be directly converted to lactate.

M3 reflects fully labelled alanine, pyruvate and lactate, whereas M0 abundances reflect metabolites with no  $^{13}\text{C}$  label. M1 and M2 represent partially labelled species that are not present in substantial amounts. **d**, Atom-transition map depicting a model for the entry of amino-acid-derived carbons into the TCA cycle and isotopic labelling of various metabolites. Open circles represent unlabelled carbon, and different-coloured circles highlight labelling patterns that correspond to specific pathways as indicated in key. For all graphs, error bars indicate mean  $\pm$  s.d. for three independent experiments. IDH, isocitrate dehydrogenase; MID, mass isotopomer distribution.

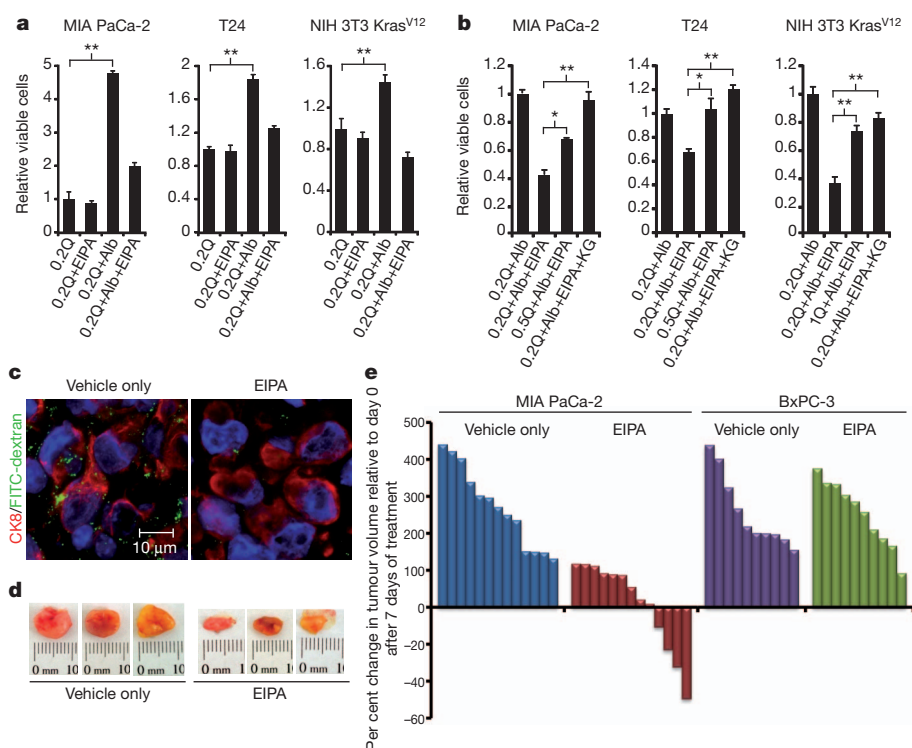
NIH 3T3 Src<sup>Y527F</sup> cells displayed increased levels of macropinocytosis relative to untransformed control cells (Supplementary Fig. 13a). Moreover, Src-induced macropinosomes displayed the capacity to internalize and degrade extracellular albumin (Supplementary Fig. 13b). Importantly, the decrease in proliferation of Src-transformed cells following glutamine deprivation (Supplementary Fig. 10) could be rescued by extracellular albumin (Supplementary Fig. 13c), suggesting that the use of albumin by cancer cells to augment amino acid supply can be mediated by inducers of macropinocytosis other than Ras.

The above observations implicate macropinocytosis in providing nutrients to sustain cancer cell proliferation; therefore we sought to directly evaluate the role of macropinocytosis in tumour growth. Mice bearing MIA PaCa-2-derived heterotopic tumours were treated with EIPA or vehicle using an osmotic pump. EIPA administration commenced when tumours attained an average volume of 50–100 mm<sup>3</sup>. After 7 days of treatment, FITC-dextran was delivered intratumorally and the tumours were excised. Tumour tissue collected from EIPA-treated animals displayed a reduction in macropinocytic uptake of FITC-dextran compared to vehicle-only controls (Fig. 4c). Moreover, relative to control tumours, those from EIPA-treated animals displayed an attenuation in growth and, in some cases regression (Fig. 4d, e). By contrast, EIPA administration had no effect on the growth rate of tumours derived from BxPC-3 cells, which display low levels of macropinocytosis (Fig. 4e). These results indicate that a reduction in macropinocytic capacity may compromise tumour growth.

The importance of albumin catabolism in human Ras-driven tumours depends on the extent to which the tumours require certain

amino acids in excess of the amounts readily attainable in free form from their environment. In human cancers, oncogenic Ras mutations are most common in adenocarcinomas of the pancreas<sup>25,26</sup>. In preliminary metabolomic experiments, we have measured the abundances of 123 known water-soluble metabolites in human pancreatic cancer tumour specimens compared to benign adjacent pancreatic tissue. The most depleted metabolite in pancreatic tumour tissue in comparison to adjacent normal tissue was glutamine, with serine and glycine also decreased (data not shown). A common feature of these amino acids is their consumption in nucleotide synthesis. These preliminary data indicate that human pancreatic tumour tissue is depleted of these free amino acids and raise the possibility that in these depleted conditions, the acquisition of amino acids through macropinocytosis and protein degradation could contribute to tumour growth. In previous experiments investigating human tumour metabolism, it was found that the production of nitrogen waste from some colorectal tumours could be in tenfold excess relative to their uptake of free amino acids<sup>27</sup>. The Ras-induced use of plasma proteins as a source of precursors for macromolecular synthesis and anaplerosis could explain both of these observations, a possibility that remains to be explored in future studies.

Recent years have witnessed a renewed appreciation of the altered metabolic behaviour of tumour cells and the critical role that such metabolic reprogramming has in conferring growth and survival advantages to tumour cells. Here we provide evidence that macropinocytosis-mediated internalization of extracellular protein and its subsequent intracellular degradation may define a mechanism for amino acid supply in Ras-transformed cancer cells. Moreover, these findings



**Figure 4 | Macropinocytosis is required for albumin-dependent cancer cell proliferation *in vitro* and for tumour growth *in vivo*.** **a**, The compromised proliferation of oncogenic Ras-expressing cells resulting from growth in media containing sub-physiological concentrations of glutamine (0.2 mM, 0.2Q) is reversed by supplementation with 2% albumin (0.2Q + Alb) and this effect is inhibited by treatment with 25 μM EIPA (0.2Q + Alb + EIPA). Total viable cell counts were measured using an MTT assay after 6 days of growth. Data are presented relative to the values obtained for the 0.2Q condition. **b**, The effects of EIPA treatment (25 μM) are suppressed by increasing the glutamine levels in the growth media to the indicated concentrations (that is, 0.5Q indicates 0.5 mM glutamine) or the addition to the medium of 7 mM dimethyl  $\alpha$ -ketoglutarate (KG). Data are presented relative to the values obtained for the

0.2Q + Alb condition. For both **a** and **b**, error bars indicate mean  $\pm$  s.e.m. for  $n = 3$  independent experiments. Statistical significance was determined using Student's  $t$ -test; \* $P < 0.05$ , \*\* $P < 0.01$ . **c–e**, EIPA inhibits macropinocytosis *in vivo* and reduces tumour growth in a subcutaneous heterotopic xenograft model of pancreatic cancer. **c**, Representative images from sections of FITC-dextran (green) injected MIA PaCa-2 tumour xenografts from mice treated with EIPA or vehicle-only controls after 7 days of treatment. The human pancreatic cancer cells are marked by anti-CK8 staining (red). **d**, Representative digital photographs of dissected tumours from mice treated with EIPA or vehicle-only controls. **e**, Waterfall plots indicating the per cent change in tumour volume after 7 days of treatment relative to baseline (day 0 of treatment) for tumours derived from MIA PaCa-2 cells or BxPC-3 cells.



raise the question of whether the inhibition of macropinocytosis can be used for therapeutic targeting in a subset of cancers.

## METHODS SUMMARY

Macropinosomes were imaged using an Axiovert 200 inverted fluorescent microscope (Zeiss) and analysed using the 'Analyze Particles' feature in ImageJ (National Institutes of Health). For the heterotopic xenografts, female homozygous NCr nude mice (Taconic) were injected subcutaneously, and when tumours reached an average volume of 500 mm<sup>3</sup> FITC-dextran (Invitrogen) was injected intratumorally. To evaluate the effects of EIPA on macropinocytosis, tumour-bearing mice were treated with EIPA (20 mg ml<sup>-1</sup>) using an osmotic pump (Alzet, Model 1004) at a pump rate of 0.11 µl h<sup>-1</sup>. Treatment was initiated when tumours attained an average volume of 50–100 mm<sup>3</sup>. As a control, we used animals treated with vehicle only (DMSO in PBS). For tumour growth assays, volumes of the subcutaneous tumours were calculated on the basis of measurements obtained using digital calipers after 7 days of treatment. For the generation of <sup>13</sup>C-labelled proteins, a prototrophic haploid SK1 strain of *Saccharomyces cerevisiae* was grown in synthetic complete media lacking amino acids (DIFCO) and containing 2% glucose that was either unlabelled, or uniformly <sup>13</sup>C-labelled (Cambridge Isotope Labs). Whole-cell extracts were prepared according to ref. 28. For intracellular metabolite analysis, cell lysate was collected and after extraction, the aqueous phase was collected and evaporated under nitrogen. Gas chromatography/mass spectrometry analysis was performed using an Agilent 6890 GC coupled to a 5975C MS. Mass isotopomer distributions were determined by integrating the appropriate ion fragments and corrected for natural isotope abundance. For glutamine-deprivation assays, cells were cultured in the indicated glutamine concentration and viable cell counts determined using the MTT assay. For some conditions, media was supplemented with BSA (Fraction V, fatty-acid-, nuclease- and protease-free (Calbiochem)). For rescue experiments, dimethyl  $\alpha$ -ketoglutarate (Sigma) was used at 7 mM<sup>29</sup>. EIPA (Invitrogen) was diluted in DMSO and used at the indicated concentrations.

**Full Methods** and any associated references are available in the online version of the paper.

Received 11 July 2012; accepted 2 April 2013.

Published online 12 May; corrected online 29 May 2013 (see full-text HTML version for details).

- Bar-Sagi, D. & Feramisco, J. R. Induction of membrane ruffling and fluid-phase pinocytosis in quiescent fibroblasts by ras proteins. *Science* **233**, 1061–1068 (1986).
- Porat-Shliom, N., Kloog, Y. & Donaldson, J. G. A unique platform for H-Ras signaling involving clathrin-independent endocytosis. *Mol. Biol. Cell* **19**, 765–775 (2008).
- Walsh, A. B. & Bar-Sagi, D. Differential activation of the Rac pathway by Ha-Ras and K-Ras. *J. Biol. Chem.* **276**, 15609–15615 (2001).
- Lopez-Crapez, E., Chypre, C., Saavedra, J., Marchand, J. & Grenier, J. Rapid and large-scale method to detect K-ras gene mutations in tumor samples. *Clin. Chem.* **43**, 936–942 (1997).
- Aoki, K., Yoshida, T., Sugimura, T. & Terada, M. Liposome-mediated *in vivo* gene transfer of antisense K-ras construct inhibits pancreatic tumor dissemination in the murine peritoneal cavity. *Cancer Res.* **55**, 3810–3816 (1995).
- Ivanov, A. I. Pharmacological inhibition of endocytic pathways: is it specific enough to be useful? *Methods Mol. Biol.* **440**, 15–33 (2008).
- West, M. A., Bretscher, M. S. & Watts, C. Distinct endocytic pathways in epidermal growth factor-stimulated human carcinoma A431 cells. *J. Cell Biol.* **109**, 2731–2739 (1989).
- Capon, D. J., Chen, E. Y., Levinson, A. D., Seeburg, P. H. & Goeddel, D. V. Complete nucleotide sequences of the T24 human bladder carcinoma oncogene and its normal homologue. *Nature* **302**, 33–37 (1983).
- Knowles, M. A. & Williamson, M. Mutation of H-ras is infrequent in bladder cancer: confirmation by single-strand conformation polymorphism analysis, designed restriction fragment length polymorphisms, and direct sequencing. *Cancer Res.* **53**, 133–139 (1993).
- Wagner, M. *et al.* A murine tumor progression model for pancreatic cancer recapitulating the genetic alterations of the human disease. *Genes Dev.* **15**, 286–293 (2001).
- Hingorani, S. R. *et al.* Trp53R172H and KrasG12D cooperate to promote chromosomal instability and widely metastatic pancreatic ductal adenocarcinoma in mice. *Cancer Cell* **7**, 469–483 (2005).
- Nolan-Steva, O. *et al.* *GLI1* is regulated through Smoothened-independent mechanisms in neoplastic pancreatic ducts and mediates PDAC cell survival and transformation. *Genes Dev.* **23**, 24–36 (2009).
- Stehle, G. *et al.* Plasma protein (albumin) catabolism by the tumor itself—implications for tumor metabolism and the genesis of cachexia. *Crit. Rev. Oncol. Hematol.* **26**, 77–100 (1997).
- Gaglio, D., Soldati, C., Vanoni, M., Alberghina, L. & Chiaradonna, F. Glutamine deprivation induces abortive S-phase rescued by deoxyribonucleotides in *k-ras* transformed fibroblasts. *PLoS ONE* **4**, e4715 (2009).
- Matés, J. M. *et al.* Glutamine homeostasis and mitochondrial dynamics. *Int. J. Biochem. Cell Biol.* **41**, 2051–2061 (2009).
- Reis, R. C., Sorgine, M. H. & Coelho-Sampaio, T. A novel methodology for the investigation of intracellular proteolytic processing in intact cells. *Eur. J. Cell Biol.* **75**, 192–197 (1998).
- Newsholme, E. A., Crabtree, B. & Ardawi, M. S. Glutamine metabolism in lymphocytes: its biochemical, physiological and clinical importance. *Q. J. Exp. Physiol.* **70**, 473–489 (1985).
- Wu, M. C., Arimura, G. K. & Yunis, A. A. Mechanism of sensitivity of cultured pancreatic carcinoma to asparaginase. *Int. J. Cancer* **22**, 728–733 (1978).
- Gaglio, D. *et al.* Oncogenic K-Ras decouples glucose and glutamine metabolism to support cancer cell growth. *Mol. Syst. Biol.* **7**, 523 (2011).
- Metallo, C. M. *et al.* Reductive glutamine metabolism by IDH1 mediates lipogenesis under hypoxia. *Nature* **481**, 380–384 (2012).
- Hirai, H. & Varmus, H. E. SH2 mutants of c-src that are host-dependent for transformation are transdominant inhibitors of mouse-cell transformation by activated c-src. *Genes Dev.* **4**, 2342–2352 (1990).
- Kasahara, K. *et al.* Role of Src-family kinases in formation and trafficking of macropinosomes. *J. Cell. Physiol.* **211**, 220–232 (2007).
- Mettlen, M. *et al.* Src triggers circular ruffling and macropinocytosis at the apical surface of polarized MDCK cells. *Traffic* **7**, 589–603 (2006).
- Veithen, A., Cuppers, P., Baudhuin, P. & Courtoy, P. J. v-Src induces constitutive macropinocytosis in rat fibroblasts. *J. Cell Sci.* **109**, 2005–2012 (1996).
- Bos, J. L. ras oncogenes in human cancer: a review. *Cancer Res.* **49**, 4682–4689 (1989).
- Jones, S. *et al.* Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* **321**, 1801–1806 (2008).
- Holm, E. *et al.* Substrate balances across colonic carcinomas in humans. *Cancer Res.* **55**, 1373–1378 (1995).
- Tsakraklides, V. & Bell, S. P. Dynamics of pre-replicative complex assembly. *J. Biol. Chem.* **285**, 9437–9443 (2010).
- Wise, D. R. *et al.* Myc regulates a transcriptional program that stimulates mitochondrial glutaminolysis and leads to glutamine addiction. *Proc. Natl Acad. Sci. USA* **105**, 18782–18787 (2008).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We are grateful to members of the Bar-Sagi laboratory for their comments and discussions, and N. Fehrenbacher and M. Philips for sharing cell lines. This work was supported National Institutes of Health (NIH) grant R01CA055360 to D.B.-S. C.C. was supported by a Canadian Institutes of Health Research postdoctoral fellowship and an AACR postdoctoral fellowship provided by the Pancreatic Cancer Action Network. M.G.V.H. acknowledges support from the Burroughs Wellcome Fund, the Damon Runyon Cancer Research Foundation, the Smith Family, the Stern family, the Broad Institute and the National Cancer Institute (P01-CA117969 and P30-CA14051-39). J.J.K. was supported by a Hope Funds for Cancer Research Fellowship (HFCR-11-03-01). C.B.T., J.A.D. and J.D.R. acknowledge support by the Stand Up To Cancer (SU2C) Pancreatic Cancer Dream Team Award. All animal care and procedures were approved by the Institutional Animal Care and Use Committee at NYU School of Medicine. The Histopathology Core of NYU School of Medicine is partially supported by the National Institutes of Health (grant 5 P30CA016087-32). Troma I, an antibody that recognizes CK8, was contributed by P. Brulet and R. Kemler and made available by the Developmental Studies Hybridoma Bank under the auspices of the NICHD.

**Author Contributions** C.C. and D.B.-S. conceived the cell biological and cell-growth experiments. C.C. carried out the macropinocytosis assays, microscopy and proliferation assays. C.C. and R.G.S.-A. carried out the xenograft experiments. C.C. and E.G. carried out the KRAS knockdown experiments. S.M.D., S.J.P., C.M.M. and M.G.V.H. conceived and carried out the <sup>13</sup>C-labelling experiments. J.J.K., S.H., M.N., J.A.D., C.B.T. and J.D.R. conceived and carried out the human metabolomics analysis.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.B.-S. ([dafna.bar-sagi@nyumc.org](mailto:dafna.bar-sagi@nyumc.org)).

## METHODS

**Cell culture.** All cells were maintained under 5% CO<sub>2</sub> at 37 °C in medium supplemented with 10% FBS (Gibco). MIA PaCa-2 and T24 cells were maintained in DMEM (Invitrogen); BxPC-3 and 5637 cells were maintained in RPMI (Gibco) supplemented with 1 mM sodium pyruvate (CellGro) and NIH 3T3 cells were maintained in DMEM supplemented with 1× MEM non-essential amino acids (Sigma).

**Macropinosome visualization and quantification.** Cells were seeded onto glass coverslips. Twenty-four to forty-eight hours after cell seeding, cells were serum starved for 18 h. Macropinosomes were marked using a high-molecular-mass TMR-dextran (Invitrogen)-uptake assay wherein TMR-dextran was added to serum-free medium at a final concentration of 1 mg ml<sup>-1</sup> for 30 min at 37 °C. At the end of the incubation period, cells were rinsed five times in cold PBS and immediately fixed in 3.7% formaldehyde. Cells were DAPI-treated to stain nuclei and coverslips mounted onto slides using DAKO Mounting Media (DAKO). Images were captured using an Axiovert 200 inverted fluorescent microscope (Zeiss) and analysed using the 'Analyze Particles' feature in ImageJ (National Institutes of Health). The total particle area per cell was determined from at least five fields that were randomly selected from different regions across the entirety of each sample.

**Interference RNA-mediated knockdown of KRAS.** KRAS knockdown was achieved through the pTRIPZ doxycycline-inducible short hairpin RNA (shRNA) expression lentiviral system (Open Biosystems). shRNA sequences for scramble shRNA (5'-GGAAGTGCATTATTCTATTA-3') and KRAS shRNA 1 (5'-TAGTTGGAGCTTGTGGCGTAG-3') were cloned into pTRIPZ according to the manufacturer's protocol. The KRAS shRNA 2 construct was purchased from Open Biosystems. MIA PaCa-2 cells were transduced with lentiviral particles containing the indicated pTRIPZ scramble shRNA or KRAS shRNAs and selected with 2 µg ml<sup>-1</sup> puromycin (Calbiochem) for 3 days. shRNA expression was induced with 1 µg ml<sup>-1</sup> doxycycline and efficient knockdown was confirmed by immunoblotting with KRAS-specific antibodies (sc-30, Santa Cruz Biotechnology).

**Mice.** For the heterotopic xenografts, female homozygous NCr nude mice (Taconic) were injected subcutaneously in both flanks at 8 weeks of age with 1 × 10<sup>6</sup> cells mixed at a 1:1 dilution with BD Matrigel (BD Biosciences) in a total volume of 100 µl. When tumours reached an average volume of 500 mm<sup>3</sup>, 1 mg of fixable FITC-dextran (Invitrogen) diluted in PBS to a volume of 100 µl was injected intratumorally. At 2 h post-injection, tumours were removed and rapidly frozen in tissue-freezing medium. Experimental cohorts of at least three mice per pancreatic cancer cell line, each mouse implanted in both flanks, were used in triplicate experiments. To quantify FITC-dextran uptake in the tumours, the total particle area per cell was determined from at least five sections per tumour with five fields analysed per section, totalling at least 15 randomly selected fields per tumour.

For the autochthonous model, pancreata from *P48-cre; Isl-Kras<sup>G12D</sup>; Trp53<sup>-/-</sup>* (KPC) mice, at 12 weeks of age, were injected with 2 mg of FITC-dextran and subsequently collected. As a control, we analysed pancreatic tissue from wild-type mice. Frozen sections from sectioned pancreata were analysed using standard microscopic techniques and macropinosocytosis-positive cells were identified by the visualization of FITC-positive puncta.

To evaluate the effects of EIPA on macropinosocytosis *in vivo*, mice heterotopically transplanted with MIA PaCa-2 cells, as described above, were treated with an osmotic pump (Alzet) when tumours attained an average volume of 50–100 mm<sup>3</sup> (approximately 2 weeks after transplantation). EIPA (20 mg ml<sup>-1</sup> in a 1:4 dilution of DMSO and PBS) was administered at a pump rate of 0.11 µl h<sup>-1</sup>. As a control, we used animals treated with vehicle only (DMSO in PBS). Animals were intratumorally injected with FITC-dextran at various time points (4–7 days) during the course of treatment. After a subsequent 1-h time interval, tumours were collected, sectioned and analysed using standard microscopic techniques. Each experimental group consisted of five animals and at least five sections per pancreas were analysed.

For tumour growth assays, the animals were randomized into control and treated groups. Treatment, as described above, was initiated 2 weeks post-transplantation when the tumours had attained an average volume of 50–100 mm<sup>3</sup>. Volumes of the subcutaneous tumours were calculated on the basis of measurements obtained in a blinded fashion using digital calipers after 7 days of treatment.

**Metabolite quantification.** NIH 3T3 Kras<sup>V12</sup> cells were grown in DMEM with/without the supplementation of 2% BSA for 24 h before collection. Cells were lysed

in a buffered 1% Triton solution and lysates were de-proteinized using perchloric acid-precipitation (BioVision). Relative metabolite concentrations were then determined using colorimetric assay kits: glutamate/glutamine (Bioassay Systems) and α-ketoglutarate (BioVision).

**Generation of <sup>13</sup>C-labelled proteins.** A prototrophic haploid SK1 strain of *Saccharomyces cerevisiae* was grown to an attenuation of 4.0 in synthetic complete media lacking amino acids (DIFCO) and containing 2% glucose that was either normal isotopic, or uniformly <sup>13</sup>C-labelled (Cambridge Isotopes). Whole-cell extracts were prepared according to ref. 28. In brief, cells were collected, re-suspended in 100 mM HEPES, pH 7.6, 0.8 mM sorbitol, 10 mM magnesium acetate, 2 mM EDTA and 300 mM potassium glutamate and lysed using a Sample Prep 6870 Freezer/Mill (SPEX CertiPrep). Lysate was treated with DNase (Ambion) and RNaseH (Invitrogen) and dialysed against PBS. Endogenous protease and glutaminase activity were eliminated by adding the protein to FBS and incubating for 30 min at 56 °C. The resulting labelled and unlabelled protein was either: (1) tagged with N-hydroxysuccinimide-rhodamine (Pierce) according to manufacturer's instructions for visualization in confocal microscopy experiments, or (2) used to supplement cell culture media for subsequent tracing experiments.

**Metabolite extraction.** For extracellular metabolite analysis, media was collected and metabolites extracted using 0.3 ml ice-cold acetone containing 1 µg norvaline. Extracts clarified by centrifugation and the supernatant evaporated under nitrogen. For intracellular metabolite analysis, cells were rinsed three times with cold saline and quenched with cold methanol. One half volume of cold water containing 1 µg norvaline was added, cell lysate was collected, and one volume of chloroform was added to each sample. After extraction the aqueous phase was collected and evaporated under nitrogen.

**Gas chromatography/mass spectrometry (GC/MS) analysis.** Dried polar metabolites were dissolved in 20 µl of 2% methoxyamine hydrochloride in pyridine (Thermo) and held at 37 °C for 1.5 h. After dissolution and reaction, tert-butyldimethylsilyl derivatization was initiated by adding 30 µl *N*-methyl-*N*-(tert-butyldimethylsilyl)trifluoroacetamide + 1% tert-butyldimethylchlorosilane (Regis) and incubating at 37 °C for 1 h. GC/MS analysis was performed using an Agilent 6890 GC equipped with a 30m DB-35MS capillary column connected to an Agilent 5975B MS operating under electron impact ionization at 70 eV. One microlitre of sample was injected in splitless mode at 270 °C, using helium as the carrier gas at a flow rate of 1 ml min<sup>-1</sup>. For measurement of amino acids, the GC oven temperature was held at 100 °C for 3 min and increased to 300 °C at 3.5 °C min<sup>-1</sup>. The MS source and quadrupole were held at 230 °C and 150 °C, respectively, and the detector was run in scanning mode, recording ion abundance in the range of 100–605 *m/z*. MID's were determined by integrating the appropriate ion fragments<sup>30</sup> listed in Supplementary Table 1 and corrected for natural isotope abundance using in house algorithms adapted from ref. 31.

**Glutamine-deprivation assays.** Cells were plated in 96-well plate format at a density of 1,000–1,500 cells per well. Thirty-two hours after seeding, cells were rinsed briefly in PBS and incubated in the indicated glutamine-deprivation media. Glutamine-free DMEM media was supplemented to the indicated concentration of glutamine in the presence of 10% dialysed FBS and 25 mM HEPES. For some conditions, media was supplemented with BSA (Fraction V, fatty-acid-, nuclease- and protease-free, Calbiochem). The nominal 0.2% albumin concentration inherently present in complete media was supplemented such that the final concentration of BSA in the media was 2%. For rescue experiments, dimethyl α-ketoglutarate (Sigma) was used at 7 mM<sup>29</sup>. For all experiments, media was replaced every 24 h. Viable cell counts were obtained using the MTT assay and relative cell number was determined by Syto 60 staining. EIPA (Invitrogen) was diluted in DMSO and used at the indicated concentrations with vehicle-only controls consisting of DMSO alone. Images depicting cell number in 96-well format were obtained by staining with Syto 60 (Molecular Probes), a red fluorescent nucleic acid stain. Plates were then scanned with an Odyssey Imager (LI-COR) at 700 nm.

30. Antoniewicz, M. R., Kelleher, J. K. & Stephanopoulos, G. Accurate assessment of amino acid mass isotopomer distributions for metabolic flux analysis. *Anal. Chem.* **79**, 7554–7559 (2007).
31. Fernandez, C. A., Des Rosiers, C., Previs, S. F., David, F. & Brunengraber, H. Correction of <sup>13</sup>C mass isotopomer distributions for natural stable isotope abundance. *J. Mass Spectrom.* **31**, 255–262 (1996).

# Small molecule inhibition of the KRAS–PDE $\delta$ interaction impairs oncogenic KRAS signalling

Gunther Zimmermann<sup>1\*</sup>, Björn Papke<sup>2\*</sup>, Shehab Ismail<sup>3\*</sup>, Nachiket Vartak<sup>2</sup>, Anchal Chandra<sup>2</sup>, Maïke Hoffmann<sup>4</sup>, Stephan A. Hahn<sup>4</sup>, Gemma Triola<sup>1</sup>, Alfred Wittinghofer<sup>3</sup>, Philippe I. H. Bastiaens<sup>2,5</sup> & Herbert Waldmann<sup>1,5</sup>

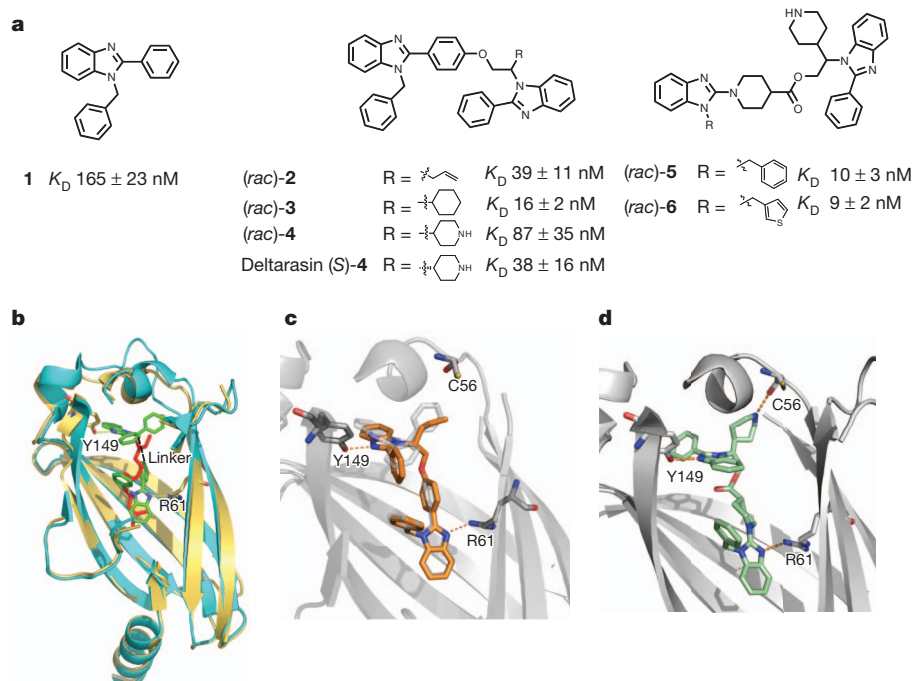
The KRAS oncogene product is considered a major target in anti-cancer drug discovery<sup>1–3</sup>. However, direct interference with KRAS signalling has not yet led to clinically useful drugs<sup>3–8</sup>. Correct localization and signalling by farnesylated KRAS is regulated by the prenyl-binding protein PDE $\delta$ , which sustains the spatial organization of KRAS by facilitating its diffusion in the cytoplasm<sup>9–11</sup>. Here we report that interfering with binding of mammalian PDE $\delta$  to KRAS by means of small molecules provides a novel opportunity to suppress oncogenic RAS signalling by altering its localization to endomembranes. Biochemical screening and subsequent structure-based hit optimization yielded inhibitors of the KRAS–PDE $\delta$  interaction that selectively bind to the prenyl-binding pocket of PDE $\delta$  with nanomolar affinity, inhibit oncogenic RAS signalling and suppress *in vitro* and *in vivo* proliferation of human pancreatic ductal adenocarcinoma cells that are dependent on oncogenic KRAS. Our findings may inspire novel drug discovery efforts aimed at the development of drugs targeting oncogenic RAS.

The interaction between a biotinylated and farnesylated KRAS4B peptide<sup>12</sup> with His-tagged PDE $\delta$  was used in an high-throughput Alpha Screen (Supplementary Fig. 1) to identify small molecules that bind to the farnesyl-binding pocket of PDE $\delta$ . The screen yielded several

benzimidazole hits (for example, **1**, Fig. 1a) which were further characterized by means of a fluorescence polarization assay based on a known PDE $\delta$  ligand<sup>13</sup> ( $K_D = 165 \pm 23$  nM for compound **1**), isothermal titration calorimetry ( $K_D = 217 \pm 15$  nM for **1**) and change in protein melting temperature upon interaction (see Supplementary Figs 2, 3 and 4)<sup>14</sup>.

Crystal structure analysis of **1** in complex with PDE $\delta$  at 1.87 Å resolution (Fig. 1b and Supplementary Fig. 5) and comparison with the previously obtained structure of the complex between PDE $\delta$  and farnesylated RHEB (root mean squared deviation (r.m.s.d.) of 0.9 Å), revealed that two benzimidazoles bind into the hydrophobic tunnel in PDE $\delta$  (Fig. 1b). One molecule is deeply buried and overlaps with the farnesyl-binding site. The second molecule is located in the vicinity of the binding site that makes main chain contacts with two carboxy-terminal RHEB amino acids. Binding of the inhibitors is mediated by hydrophobic interactions and hydrogen bonding between the benzimidazole units and Tyr 149 and Arg 61 (Fig. 1b). Consistently, related N-benzylated 2-phenylindole shows no binding (Supplementary Fig. 3). Furthermore, the side chain of Trp 90 underwent a conformational change resulting in a T stacking with inhibitor **1** (Supplementary Fig. 6).

Structure-guided design based on the crystal structure obtained for **1** in complex with PDE $\delta$  suggested covalent linking of the benzimidazoles



**Figure 1 | Structure-based development of inhibitors.**

**a**, Structure and binding affinities of benzimidazole compounds **1–6** as determined by competitive fluorescence polarization assay (see Supplementary Information, Supplementary Fig. 3). **b**, Ribbon diagram of PDE $\delta$  structure in complex with **1** (yellow), and overlay with the previously obtained crystal structure of farnesylated RHEB peptide with PDE $\delta$  (cyan). Small molecule **1** (green) and farnesyl group (red) are shown as ball and sticks. Hydrogen bonding interactions between two molecules of **1** and Tyr 149 and Arg 61 in the co-crystal structure are highlighted. **c**, Structure of **2** (orange sticks) in complex with PDE $\delta$ . Overlaid is the structure of two molecules of **1** (faint grey sticks) in complex with PDE $\delta$ . **d**, Crystal structure of (S)-**5** in complex with PDE $\delta$  confirms the presence of a hydrogen bond between the piperidine and the backbone carbonyl of Cys 56.

<sup>1</sup>Department of Chemical Biology, Max Planck Institute of Molecular Physiology, D-44227 Dortmund, Germany. <sup>2</sup>Department of Systemic Cell Biology, Max Planck Institute of Molecular Physiology, D-44227 Dortmund, Germany. <sup>3</sup>Structural Biology Group, Max Planck Institute for Molecular Physiology, D-44227 Dortmund, Germany. <sup>4</sup>Department of Molecular Gastrointestinal Oncology, Ruhr-University Bochum, D-44801 Bochum, Germany. <sup>5</sup>Chemical Biology, Faculty of Chemistry, TU Dortmund, D-44227 Dortmund, Germany.

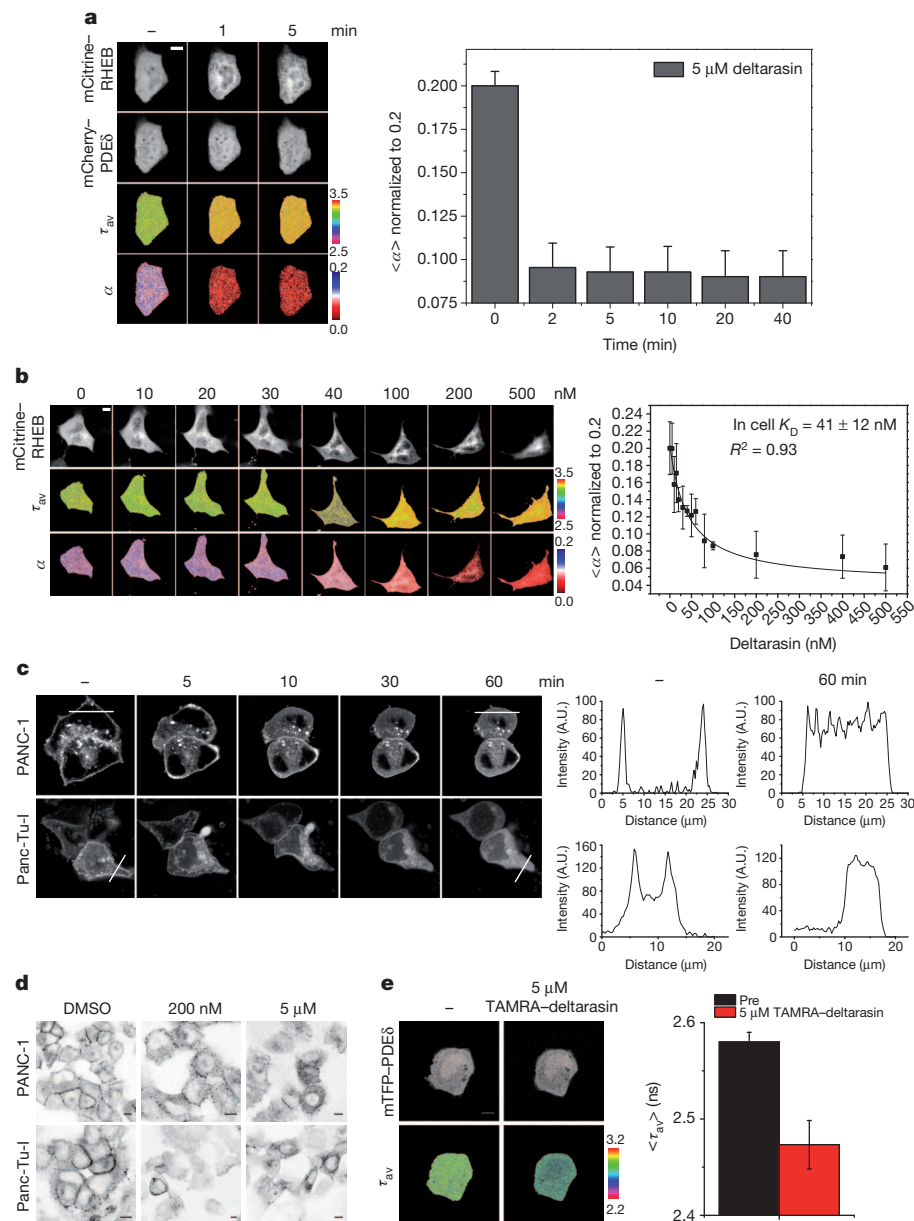
\*These authors contributed equally to this work.



**Figure 2 | In-cell measurements of the effect of deltarasin on the interaction of RAS with PDE $\delta$  and resulting delocalization of KRAS.** **a**, FLIM time series on MDCK cells expressing mCitrine–RHEB and mCherry–PDE $\delta$  show a loss of interaction between RHEB and PDE $\delta$  after treatment with 5  $\mu$ M deltarasin. Left panel, representative sample of FLIM time series. Upper

two rows show fluorescence intensity distribution of the indicated fluorescent fusion proteins, whereas the lower two rows show maps of average fluorescence lifetime ( $\tau_{av}$ ) in ns and computed molar fraction ( $\alpha$ ) of interacting mCitrine–RHEB with mCherry–PDE $\delta$ . Time in minutes is indicated above the panels. The deltarasin-induced dissociation of mCherry–PDE $\delta$  and mCitrine–RHEB is represented in the time course of normalized average  $\langle\alpha\rangle \pm$  s.e.m. for  $N = 5$  cells in the right panel. **b**, Left panel, deltarasin dose dependence of molar fraction ( $\alpha$ ) of interacting mCitrine–RHEB with mCherry–PDE $\delta$ . Upper row shows fluorescence intensity distribution of mCitrine–RHEB, middle row shows average fluorescence lifetime ( $\tau_{av}$ ) in ns and lower row shows molar fraction ( $\alpha$ ) of interacting mCitrine–RHEB with mCherry–PDE $\delta$ . The concentration of deltarasin is indicated at the top of the panel in nM. Right panel, fit of averaged dose-response  $\pm$  s.e.m. of four independent experiments to a binding model (see methods) yielded an in cell  $K_D$  of  $41 \pm 12$  nM for deltarasin binding to PDE $\delta$ .

**c**, Time series of mCitrine–KRAS redistribution upon application of 5  $\mu$ M of deltarasin in PANC-1 (upper panel) and Panc-Tu-I (lower panel) cells. Time in minutes is indicated above the panel. The first and last time point of each cell line were used to quantify the mCitrine–KRAS distribution in these cells. The loss of plasma membrane localization can be seen in the mCitrine–KRAS intensity profiles along the white lines in the fluorescence micrographs in the right panels. A.U., arbitrary units. **d**, Immunofluorescence staining of fixed and permeabilized PDAC cells with a pan antibody against RAS (Calbiochem, Anti-Pan-RAS), 2 h after administration of the vehicle DMSO, 200 nM and 5  $\mu$ M of deltarasin. **e**, FRET-FLIM measurements of the interaction between mTFP–PDE $\delta$  and TAMRA–deltarasin. Upper row, fluorescence intensity of mTFP–PDE $\delta$ ; lower row, average fluorescence lifetime ( $\tau_{av}$ ) map of mTFP–PDE $\delta$  alone and in complex with the TAMRA–deltarasin. The averaged drop in the fluorescence lifetime ( $\langle\tau_{av}\rangle \pm$  s.d.) of mTFP–PDE $\delta$  due to FRET with TAMRA–deltarasin is presented in the bar graph at the right side for  $N = 3$  in each condition. Scale bars in all micrographs indicate 10  $\mu$ m.



and synthesis of a focused library (Fig. 1a) with variation of the linker structure (see the Supplementary Information). Covalent linkage via an ether bond for instance yielded dimeric compound **2** (Fig. 1a), which binds to PDE $\delta$  with significantly increased affinity ( $K_D = 39 \pm 11$  nM) and with an almost complete overlap with the positions of the individual benzimidazoles (Fig. 1c).

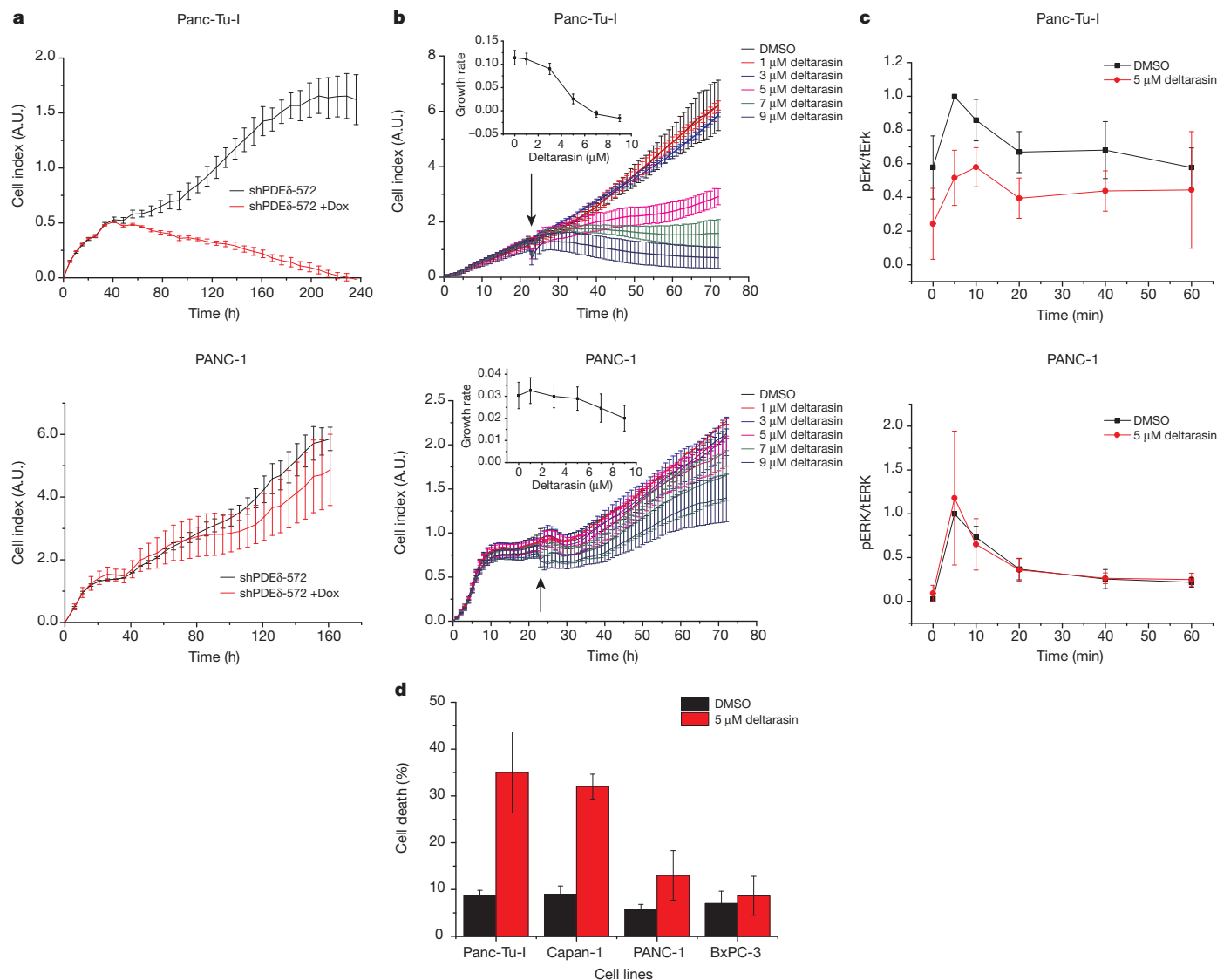
Investigation of the crystal structure of the complex (Fig. 1c) indicated replacement of the allyl group at R for a larger cyclohexyl moiety, which led to increased affinity (**3**,  $K_D = 16 \pm 2$  nM, Fig. 1a), whereas introduction of a negative charge ( $R = \text{CH}_2\text{COOH}$ ,  $K_D = 870 \pm 290$  nM, see also Supplementary Table 1), omission of a substituent ( $R = \text{H}$ ,  $K_D = 116 \pm 29$  nM) or increased steric bulk ( $R = \text{Boc}(4\text{-piperidine})$ ,  $K_D > 2000$  nM), yielded less potent ligands.

The proximity of the backbone carbonyl of Cys 56 to substituent R (Fig. 1c) suggested introduction of a piperidine to introduce an

additional hydrogen bond (Fig. 1c). However, piperidine-containing compound **4** (Fig. 1a) had lower affinity compared to **3**, probably due to rigidity of the linker that might not allow for three hydrogen-bonding interactions.

Replacement of the phenyl ether by a flexible piperidine 4-carboxylic acid ester resulted in an improvement of  $K_D$  by almost one order of magnitude (**5**,  $K_D = 10 \pm 3$  nM, Fig. 1a). Crystal structure analysis of **5** in complex with PDE $\delta$  confirmed the presence of a hydrogen bond between the piperidine moiety in **5** and the carbonyl backbone of Cys 56 (Fig. 1d). Replacement of the benzyl moiety with a 3-methyl thiophene lead to a similarly potent compound (**6**,  $K_D = 9 \pm 2$  nM, Fig. 1a).

Separation of the enantiomers of Boc-protected **4** and **6** by preparative chiral high-performance liquid chromatography (Supplementary Fig. 7), removal of the Boc group and biochemical investigation of the



**Figure 3 | Inhibition of PDE $\delta$ –KRAS interaction suppresses proliferation and MAPK-signalling in oncogenic KRAS-dependent PDAC cells.** **a**, Real-time cell analysis (RTCA) of oncogenic KRAS-dependent Panc-Tu-I (upper panel) and oncogenic RAS-independent PANC-1 (lower panel) PDAC cell proliferation after doxycycline-induced PDE $\delta$  knock-down. Cells were transduced with shPDE $\delta$ -572 doxycycline-inducible knockdown vector. Cell indices  $\pm$  s.d. were measured in duplicates. Cells were treated with doxycycline to induce PDE $\delta$  knockdown from the beginning of the experiment (+ Dox). **b**, RTCA of deltarasin dose PDAC cell proliferation response  $\pm$  s.d. of oncogenic KRAS-dependent (Panc-Tu-I, upper panel) and KRAS-independent (PANC-1, lower panel) cell lines shows deltarasin-induced suppression of proliferation in oncogenic KRAS-dependent Panc-Tu-I cells. Deltarasin was added at the indicated time point (arrow) and concentration (1–9  $\mu$ M). The inset in each panel shows deltarasin dose versus growth-rate response  $\pm$  s.d. as determined from the average of the first derivative of the cell growth curves

pure enantiomers revealed a four to sixfold difference in potency for each pair of enantiomers ((*S*)-**4**  $K_D$  =  $38 \pm 16$  nM vs (*R*)-**4**  $K_D$  =  $190 \pm 55$  nM), and ((*S*)-**6**  $K_D$  =  $7 \pm 3$  nM vs (*R*)-**6**  $K_D$  =  $39 \pm 18$  nM). The absolute configuration was assigned by analogy to a previously reported synthesis (see the Supplementary Information)<sup>15</sup>. Tight binding of (*S*)-**4** and (*S*)-**6** to PDE $\delta$  was verified by means of a direct titration using 5-carboxytetramethyl rhodamine (TAMRA)-labelled (*S*)-**4** ( $K_D$  =  $7.6 \pm 1.3$  nM) and TAMRA-(*S*)-**6** ( $K_D$  =  $5.3 \pm 1.5$  nM, Supplementary Fig. 8). TAMRA-(*S*)-**4** and TAMRA-(*S*)-**6** did not bind to the PDE $\delta$ -homologous GDI-like solubilizing factors HRG4/Unc119a

determined between 35 and 65 h (Supplementary Fig. 15). **c**, EGF-induced MAPK signalling response in PDAC cells treated with deltarasin. Peak normalized Erk1/Erk2 phosphorylation time profiles upon stimulation with  $200 \text{ ng ml}^{-1}$  EGF in serum starved Panc-Tu-I and PANC-1 cells as quantified from three independent western blots for each cell line. Each western blot contained the Erk1/Erk2 time response with vehicle control and deltarasin (Supplementary Fig. 18). The average  $\pm$  s.d. is shown for each time point. Black, vehicle control DMSO; red, 2-h incubation with 5  $\mu$ M deltarasin before EGF administration. **d**, Deltarasin induces cell death in KRAS-dependent PDAC cells (Panc-Tu-I, Capan-1) as measured by an annexin-V/propidium iodide FACS analysis (Supplementary Fig. 16). The bar graph shows the average  $\pm$  s.d. of three independent experiments for each cell line and condition. Cells were analysed by FACS after 24 h of vehicle DMSO (black) and 5  $\mu$ M deltarasin incubation (red).

and Unc119b<sup>16,17</sup>, nor to the presumed prenyl-binding proteins galectin-1 and galectin-3 (refs 18–20), indicating specificity for PDE $\delta$  (Supplementary Fig. 9). Compound **4** contains a hydrolytically stable ether linker (as opposed to an ester in **6**) and has appreciable solubility and membrane permeability (Supplementary Information), (*S*)-**4** was therefore used in the following cell biological studies. We term this compound deltarasin.

Fluorescence lifetime imaging microscopy (FLIM)-based fluorescence resonance energy transfer (FRET) measurements<sup>21</sup> were used to address whether deltarasin affected the interaction of PDE $\delta$  with

KRAS in live cells. To this end, mCherry-PDE $\delta$  was expressed together with mCitrine-tagged farnesylated RAS variants (KRAS6Q mutant or the RAS family protein RHEB) that lack part of the polybasic plasma membrane interaction motif of KRAS. Thus, their soluble fraction and interaction with PDE $\delta$  are strongly enhanced<sup>9</sup>. This largely facilitates the detection of the effect of small molecules on the interaction between mCitrine-tagged farnesylated RAS and mCherry-PDE $\delta$  in the cytoplasm of live cells by FLIM. The fluorescence patterns of mCitrine-RHEB or mCitrine-KRAS6Q and mCherry-PDE $\delta$  in Madine-Darby canine kidney (MDCK) cells transiently co-transfected with these proteins were homogeneous, showing a clear solubilization of mCitrine-RHEB/KRAS6Q by mCherry-PDE $\delta$ . A substantial drop in the mCitrine fluorescence lifetime showed that FRET occurred between the fluorescent proteins and corroborated that the solubilization of RHEB/KRAS6Q was due to a direct interaction with PDE $\delta$  (Fig. 2a and Supplementary Fig. 10). Computation of the fraction of interacting molecules ( $\alpha$ ) by global analysis of the fluorescence decay profiles as obtained by FLIM<sup>21–23</sup> showed that a significant fraction of mCitrine-RHEB/KRAS6Q was in complex with mCherry-PDE $\delta$ . Within a minute, 5  $\mu$ M of deltarasin completely inhibited this interaction and released the insolubilized mCitrine-RHEB/KRAS6Q to the endomembrane system (Fig. 2a and Supplementary Fig. 10). This shows that deltarasin interferes with the binding of KRAS to PDE $\delta$  in cells and thereby inhibits its solubilization. Quantification of the interaction between mCherry-PDE $\delta$  and mCitrine-RHEB (or mCitrine-KRAS6Q) in live cells by global analysis of FLIM data as function of deltarasin dose (Fig. 2b and Supplementary Fig. 11) enabled the computation of an 'in cell'  $K_D$  of  $41 \pm 12$  nM ( $27 \pm 7$  nM for the mCitrine-KRAS6Q assay) for the PDE $\delta$ -deltarasin interaction, remarkably similar to that determined for binding of deltarasin to purified PDE $\delta$  ( $38 \pm 16$  nM).

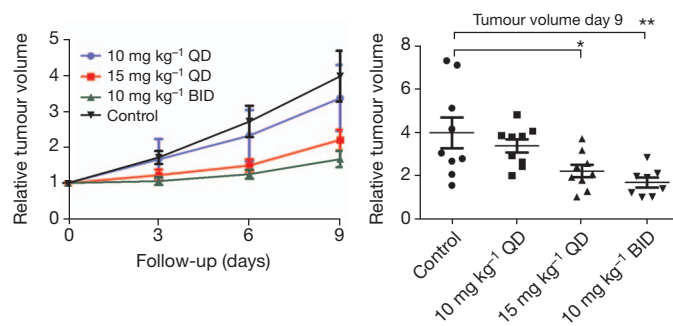
By analogy to PDE $\delta$  knockdown<sup>9</sup>, treatment of human pancreatic ductal adenocarcinoma (PDAC) cell models PANC-1 and Panc-Tu-I with 5  $\mu$ M deltarasin led within 1 h to a clear loss of mCitrine-KRAS (ectopically expressed) plasma membrane localization and its redistribution to endomembranes (Fig. 2c). Immunofluorescence staining of this type of PDAC cells with an anti-RAS antibody also showed that deltarasin induced a random distribution of endogenous RAS to endomembranes (Fig. 2d). Direct binding of deltarasin to PDE $\delta$  in live PANC-1 cells could also be demonstrated by the occurrence of FRET between monomeric teal fluorescent protein-tagged PDE $\delta$  (mTFP-PDE $\delta$ ) and TAMRA-tagged deltarasin, as seen by the significant reduction in fluorescence lifetime of mTFP in the presence of TAMRA-deltarasin (Fig. 2e). These experiments show that inhibition of the PDE $\delta$ -KRAS interaction by binding of deltarasin to PDE $\delta$  leads to a loss of KRAS spatial organization in PDAC cells, as maintained by the solubilizing activity of PDE $\delta$ .

To assess the effect of deltarasin on oncogenic KRAS signalling, the growth of PDAC cells that are dependent on oncogenic KRAS for their survival (Panc-Tu-I and Capan-1 cells) was compared to PDAC cells that are independent of oncogenic KRAS (PANC-1 cells)<sup>24</sup> or express wild-type KRAS (BxPC-3 cells)<sup>25</sup>. Transfection of these PDAC cells with a lentiviral construct for doxycycline-inducible short hairpin RNA (shRNA) expression against PDE $\delta$  (Supplementary Figs 12 and 13) and treatment with doxycycline resulted in strongly reduced cell proliferation and cell death of the KRAS-dependent Panc-Tu-I and Capan-1 PDAC cell lines after 2–4 days and had only a slight effect on the growth of the other cell lines (Fig. 3a and Supplementary Fig. 14a). A similar effect on the proliferation of KRAS-dependent PDAC cells was found with the PDE $\delta$  inhibitor deltarasin (Fig. 3b and Supplementary Fig. 14b). The PDAC proliferation profiles as function of deltarasin dose (1–9  $\mu$ M) show that the KRAS-dependent Panc-Tu-I cells show strongly reduced proliferation and increased cell death at around 5  $\mu$ M deltarasin, whereas KRAS-independent PANC-1 cells only have a small negative dependence of growth rate on deltarasin concentration (Fig. 3b, d and Supplementary Figs 15 and 16). The

KRAS-dependent Capan-1 cells also showed strongly reduced proliferation and increased cell death within hours upon treatment with 5  $\mu$ M deltarasin, whereas the wild-type KRAS-harbouring BxPC-3 PDAC cells only had a slightly reduced initial growth rate and little or no increase in cell death with respect to the control (Fig. 3d and Supplementary Figs 14 and 16). These experiments indicate that the reduced proliferation of KRAS-dependent PDAC cells is caused in part by attenuated survival signalling from oncogenic KRAS that is delocalized on endomembranes. The deltarasin concentration ( $\sim 3$   $\mu$ M, Fig. 3b) that induced a measurable effect on the proliferation of Panc-Tu-I cells was higher than that of complete inhibition of the interaction between PDE $\delta$  and RAS ( $\sim 200$  nM, Fig. 2b). Because deltarasin was found to be stable to metabolism by PDAC cells (Supplementary Fig. 17), this is probably due to the activity of ABC-transporters<sup>26</sup> that oppose the inward flux of the compound and thereby lower its intracellular availability at longer times.

To address whether KRAS relocalization by deltarasin-mediated PDE $\delta$  inhibition has an effect on oncogenic KRAS signal transduction by uncoupling it from its effectors at the plasma membrane<sup>9</sup>, the epidermal growth factor (EGF)-induced mitogen-activated protein kinase (MAPK) response in serum-starved PDAC cells was studied. Quantitative western blot analysis of the phosphorylated Erk1 and Erk2 MAPKs showed that the KRAS-dependent Panc-Tu-I cells had a high basal Erk activity that was reduced upon inhibition of PDE $\delta$  by deltarasin (Fig. 3c and Supplementary Fig. 18). Strikingly, the two oncogenic KRAS-dependent PDAC cell lines (Panc-Tu-I and Capan-1) showed a substantial reduction in the EGF-mediated transient MAPK signal response as compared to the other cell lines (Fig. 3c and Supplementary Figs 14 and 18). The residual transient MAPK response to EGF can probably be attributed to the remaining wild-type RAS isoforms at the plasma membrane that are encoded by the healthy (wild-type) alleles. Therefore, oncogenic KRAS-dependent proliferative and survival signalling is strongly attenuated by the loss of KRAS plasma membrane localization as caused by the inhibition of PDE $\delta$ -KRAS interaction by deltarasin.

To evaluate the anti-tumour activity of deltarasin, we administered the compound to nude mice bearing subcutaneous human Panc-Tu-I tumour cell xenografts and monitored tumour growth rate. Deltarasin was administered by intra-peritoneal (i.p.) injection once (QD) or twice (BID) per day. Three dosage regimens were tested (10 mg kg<sup>-1</sup> QD, 15 mg kg<sup>-1</sup> QD and 10 mg kg<sup>-1</sup> BID). An initial  $\sim 15\%$  maximal weight loss of mice injected with deltarasin could be observed during the first 2 days of treatment, after which this stabilized in all groups (Supplementary Fig. 19). A clear dose-dependent reduction in Panc-Tu-I



**Figure 4 | Deltarasin impairs dose-dependent *in vivo* growth of xenografted pancreatic carcinoma in nude mice.** a, b, Tumour volume measurements (a) and tumour volume distribution (b) at day 9 of Panc-Tu-I xenograft tumours treated with vehicle or deltarasin at the dosages indicated: deltarasin was administered by intra-peritoneal injection once (QD) or twice (BID) per day at 10 mg kg<sup>-1</sup> QD, 15 mg kg<sup>-1</sup> QD and 10 mg kg<sup>-1</sup> BID. Changes in mean tumour volumes are given relative to the volumes at treatment initiation. Error bars represent s.e.m. with  $n = 9$  for the control, 10 mg kg<sup>-1</sup> QD, and 15 mg kg<sup>-1</sup> QD and  $n = 8$  for the 10 mg kg<sup>-1</sup> BID group.  $P$  values were obtained by unpaired  $t$ -test. \*\* $P \leq 0.01$ , \* $P \leq 0.05$ .



tumour growth rate could be observed in deltarasin treated mice with respect to the vehicle-injected controls, where the growth of tumours in mice that were treated with 10 mg kg<sup>-1</sup> BID deltarasin was almost completely blocked (Fig. 4a). The negative effect of the compound on Panc-Tu-I tumour growth could also be observed in the reduced variance in tumour size with respect to the control as measured at day 9 (Fig. 4b).

Our results demonstrate that inhibition of the PDEδ–KRAS interaction by means of small molecules affects the spatial organization of KRAS and thus provides a novel opportunity to suppress oncogenic RAS signalling and thereby tumour growth.

## METHODS SUMMARY

Screening based on Alpha-technology was conducted in white, non-binding 1,536-well plates in a final volume of 6 µl (His<sub>6</sub>-PDEδ, 100 nM, biotinylated KRAS peptide 250 nM).

K<sub>D</sub> values were measured by fluorescence polarization measurements. For direct titrations, increasing amounts of PDEδ were added to a solution containing 50–100 nM labelled small molecule in 200 µl PBS buffer. For displacement titrations, increasing amounts of the small molecules in DMSO were directly added to fluorescein-labelled atorvastatin (24 nM) and His<sub>6</sub>-tagged PDEδ (40 nM) in 200 µl PBS-buffer (containing 0.05% CHAPS, 1% DMSO), keeping the concentration of fluorescein-labelled atorvastatin, PDEδ and DMSO constant.

For K<sub>D</sub> measurements using isothermal titration calorimetry, PDEδ protein (280 µM) was titrated to small molecule (30 µM) in Tris/HCl buffer (temperature 25 °C). In the T<sub>m</sub> shift assays, protein melting points were detected by circular dichroism spectroscopy in the presence of small molecules.

Inhibitors were co-crystallized with PDEδ by mixing a solution of small molecule and PDEδ with DMSO 1.7% as a final concentration. Crystals were obtained from a Qiagen crystallization screen.

For live-cell microscopy, cells were grown in four-well Lab-Tek chambers (NUNC) and transferred to low-bicarbonate DMEM without phenol red supplemented with 25 mM HEPES at pH 7.4.

The following antibodies were used for western blotting: total Erk (Abcam-AB36991; 1:2,000), pErk (Cell Signaling-9101; 1:2,000) and infrared secondary antibodies (LI-COR).

Fluorescence lifetime images were acquired using a confocal laser-scanning microscope (FV1000, Olympus) equipped with a time-correlated single-photon counting module (LSM Upgrade Kit, Picoquant). Intensity thresholds were applied to segment the cells from the background fluorescence. Data were further analysed as described in ref. 23 to obtain images of the molar fraction (α) of interacting mCherry–PDEδ/mCitrine–RHEB.

Real-time cell analyses were carried out by monitoring the proliferation for at least 3 days after administration of the inhibitor.

Received 23 August 2012; accepted 19 April 2013.

Published online 22 May 2013.

- Malumbres, M. & Barbacid, M. RAS oncogenes: the first 30 years. *Nature Rev. Cancer* **3**, 459–465 (2003).
- Gelb, M. H. et al. Therapeutic intervention based on protein prenylation and associated modifications. *Nature Chem. Biol.* **2**, 518–528 (2006).
- Downward, J. Targeting RAS signalling pathways in cancer therapy. *Nature Rev. Cancer* **3**, 11–22 (2003).
- Berndt, N., Hamilton, A. D. & Sefti, S. M. Targeting protein prenylation for cancer therapy. *Nature Rev. Cancer* **11**, 775–791 (2011).
- Roberts, P. J. & Der, C. J. Targeting the Raf-MEK-ERK mitogen-activated protein kinase cascade for the treatment of cancer. *Oncogene* **26**, 3291–3310 (2007).
- Maurer, T. et al. Small-molecule ligands bind to a distinct pocket in Ras and inhibit SOS-mediated nucleotide exchange activity. *Proc. Natl Acad. Sci. USA* **109**, 5299–5304 (2012).
- Patgiri, A., Yadav, K. K., Arora, P. S. & Bar-Sagi, D. An orthosteric inhibitor of the Ras-Sos interaction. *Nature Chem. Biol.* **7**, 585–587 (2011).
- Sun, Q. et al. Discovery of small molecules that bind to K-Ras and inhibit Sos-mediated activation. *Angew. Chem. Int. Edn* **51**, 6140–6143 (2012).
- Chandra, A. et al. The GDI-like solubilizing factor PDEδ sustains the spatial organization and signalling of Ras family proteins. *Nature Cell Biol.* **14**, 329–329 (2012).

- Ismail, S. A. et al. Arl2-GTP and Arl3-GTP regulate a GDI-like transport system for farnesylated cargo. *Nature Chem. Biol.* **7**, 942–949 (2011).
- Zhang, H. et al. Photoreceptor cGMP phosphodiesterase δ subunit (PDEδ) functions as a prenyl-binding protein. *J. Biol. Chem.* **279**, 407–413 (2004).
- Chen, Y. X. et al. Synthesis of the Rheb and K-Ras4B GTPases. *Angew. Chem. Int. Edn* **49**, 6090–6095 (2010).
- Chidley, C., Haruki, H., Pedersen, M. G., Muller, E. & Johnsson, K. A yeast-based screen reveals that sulfasalazine inhibits tetrahydrobiopterin biosynthesis. *Nature Chem. Biol.* **7**, 375–383 (2011).
- Vedadi, M. et al. Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination. *Proc. Natl Acad. Sci. USA* **103**, 15835–15840 (2006).
- Monovich, L. G. et al. Discovery of potent, selective, and orally active carboxylic acid based inhibitors of matrix metalloproteinase-13. *J. Med. Chem.* **52**, 3523–3538 (2009).
- Wright, K. J. et al. An ARL3-UNC119-RP2 GTPase cycle targets myristoylated NPHP3 to the primary cilium. *Genes Dev.* **25**, 2347–2360 (2011).
- Zhang, H. et al. UNC119 is required for G protein trafficking in sensory neurons. *Nature Neurosci.* **14**, 874–880 (2011).
- Elad-Sfadia, G., Haklai, R., Balan, E. & Kloog, Y. Galectin-3 augments K-Ras activation and triggers a Ras signal that attenuates ERK but not phosphoinositide 3-kinase activity. *J. Biol. Chem.* **279**, 34922–34930 (2004).
- Paz, A., Haklai, R., Elad-Sfadia, G., Ballan, E. & Kloog, Y. Galectin-1 binds oncogenic H-Ras to mediate Ras membrane anchorage and cell transformation. *Oncogene* **20**, 7486–7493 (2001).
- Bhagatji, P., Leventis, R., Rich, R., Lin, C. J. & Silvius, J. R. Multiple cellular proteins modulate the dynamics of K-ras association with the plasma membrane. *Biophys. J.* **99**, 3327–3335 (2010).
- Wouters, F. S., Verveer, P. J. & Bastiaens, P. I. H. Imaging biochemistry inside cells. *Trends Cell Biol.* **11**, 203–211 (2001).
- Grecco, H. E. et al. In situ analysis of tyrosine phosphorylation networks by FLIM on cell arrays. *Nature Methods* **7**, 467–472 (2010).
- Grecco, H. E., Roda-Navarro, P. & Verveer, P. J. Global analysis of time correlated single photon counting FRET-FLIM data. *Opt. Express* **17**, 6493–6508 (2009).
- Moore, P. S. et al. Genetic profile of 22 pancreatic carcinoma cell lines. *Virchows Arch.* **439**, 798–802 (2001).
- Berrozpe, G., Schaeffer, J., Peinado, M. A., Real, F. X. & Perucho, M. Comparative analysis of mutations in the p53 and K-ras genes in pancreatic-cancer. *Int. J. Cancer* **58**, 185–191 (1994).
- Perez-Tomas, R. Multidrug resistance: retrospect and prospects in anti-cancer drug treatment. *Curr. Med. Chem.* **13**, 1859–1876 (2006).

Supplementary Information is available in the online version of the paper.

**Acknowledgements** The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Program (FP7/2007-2013)/ERC Grant agreement no. 268309 to H.W., and no. 268782 to A.W. The Compound Management und Screening Center (COMAS), Dortmund, Germany, is acknowledged for carrying out high-throughput screening and data analysis. G.Z. acknowledges the Fonds der Chemischen Industrie for a Kekulé Scholarship. We thank K. Michel for help with western blot analysis. We are grateful to C. Degenhart, A. Wolf, S. Baumann and A. Choidas for help with screening assay development and for the determination of solubility, membrane permeability and stability of deltarasin.

**Author Contributions** A.W., P.I.H.B. and H.W. conceived the project. H.W., G.T. and G.Z. developed the screen and chemistry to generate the PDEδ inhibitor. P.I.H.B. designed the cell biological experiments and A.W. with S.I. solved the structures by X-ray crystallography. A.W., H.W., S.I. and G.Z. designed the PDEδ structure-based inhibitor development. S.A.H. developed the inducible shRNA systems and the xenograft model. G.Z. synthesized the inhibitors and performed the biochemical and biophysical characterization experiments. B.P. performed the real-time cell analysis, FACS and western blot analysis. N.V., B.P. and A.C. performed the microscopy experiments. M.H. performed the shRNA real-time cell analysis and xenograft experiments. H.W. and P.I.H.B. wrote the manuscript with help from A.W., S.I., B.P., G.T. and G.Z.

**Author Information** The atomic coordinates of PDEδ in complex with inhibitor **1**, *rac-S1*, *rac-2* and (*S*)-**5** are deposited in the Protein Data Bank with accession numbers 4JV6, 4JV8, 4JVB and 4JVF, respectively. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.W. ([herbert.waldmann@mpi-dortmund.mpg.de](mailto:herbert.waldmann@mpi-dortmund.mpg.de)), P.I.H.B. ([philippe.bastiaens@mpi-dortmund.mpg.de](mailto:philippe.bastiaens@mpi-dortmund.mpg.de)) or A.W. ([alfred.wittinghofer@mpi-dortmund.mpg.de](mailto:alfred.wittinghofer@mpi-dortmund.mpg.de)).

# Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics

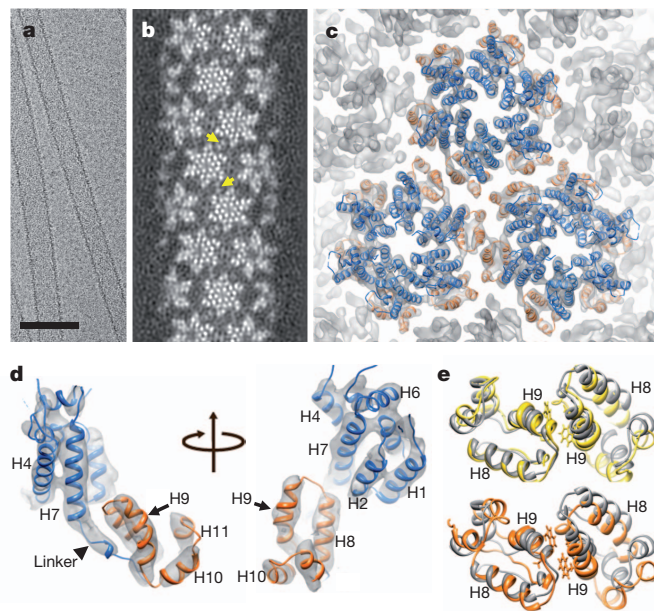
Gongpu Zhao<sup>1,2\*</sup>, Juan R. Perilla<sup>3\*</sup>, Ernest L. Yufenyuy<sup>2,4\*</sup>, Xin Meng<sup>1,2</sup>, Bo Chen<sup>5</sup>, Jiying Ning<sup>1,2</sup>, Jinwoo Ahn<sup>1,2</sup>, Angela M. Gronenborn<sup>1,2</sup>, Klaus Schulten<sup>3</sup>, Christopher Aiken<sup>2,4</sup> & Peijun Zhang<sup>1,2,6</sup>

Retroviral capsid proteins are conserved structurally but assemble into different morphologies<sup>1</sup>. The mature human immunodeficiency virus-1 (HIV-1) capsid is best described by a 'fullerene cone' model<sup>2,3</sup>, in which hexamers of the capsid protein are linked to form a hexagonal surface lattice that is closed by incorporating 12 capsid-protein pentamers. HIV-1 capsid protein contains an amino-terminal domain (NTD) comprising seven  $\alpha$ -helices and a  $\beta$ -hairpin<sup>4,5</sup>, a carboxy-terminal domain (CTD) comprising four  $\alpha$ -helices<sup>6,7</sup>, and a flexible linker with a  $3_{10}$ -helix connecting the two structural domains<sup>8</sup>. Structures of the capsid-protein assembly units have been determined by X-ray crystallography<sup>9,10</sup>; however, structural information regarding the assembled capsid and the contacts between the assembly units is incomplete. Here we report the cryo-electron microscopy structure of a tubular HIV-1 capsid-protein assembly at 8 Å resolution and the three-dimensional structure of a native HIV-1 core by cryo-electron tomography. The structure of the tubular assembly shows, at the three-fold interface<sup>11</sup>, a three-helix bundle with critical hydrophobic interactions. Mutagenesis studies confirm that hydrophobic residues in the centre of the three-helix bundle are crucial for capsid assembly and stability, and for viral infectivity. The cryo-electron-microscopy structures enable modelling by large-scale molecular dynamics simulation, resulting in all-atom models for the hexamer-of-hexamer and pentamer-of-hexamer elements as well as for the entire capsid. Incorporation of pentamers results in closer trimer contacts and induces acute surface curvature. The complete atomic HIV-1 capsid model provides a platform for further studies of capsid function and for targeted pharmacological intervention.

HIV-1 capsid protein (CA) spontaneously assembles *in vitro* into helical tubes and cones that resemble authentic viral capsids<sup>3</sup>, allowing direct structural investigation of the CA assembly using cryo-electron microscopy (cryo-EM). Previous studies of helical tubes provided low-resolution views of the curved hexagonal lattice, which is stabilized by important inter-hexamer CTD-CTD contacts across local two- and three-fold axes<sup>3,11</sup>. However, accurate and reliable atomic modelling of the HIV-1 capsid assembly has not been possible, owing to limited resolution (up to 16 Å). In fact, high-resolution structural determination of *in vitro* assembled samples is very challenging: HIV-1 capsid assembly *in vitro* is inefficient, requires high salt (1 M), and exhibits great variability in tube diameters and helical symmetries (Supplementary Fig. 1a)<sup>3,11</sup>. We devised a rapid back-dilution method to improve cryo-EM sample quality and recorded more than 1,500 cryo-EM images on film, many of which showed diffraction layer lines to 10 Å resolution. Using 27 high-resolution tubes in a single helical symmetry ( $-12, 11$ ), along with a real-space, single-particle, helical reconstruction method<sup>12,13</sup>, we determined the structure of the CA tubular assembly at 8.6 Å resolution (Fig. 1a–c and Supplementary Fig. 1b–d).

The density map clearly delineates all the  $\alpha$ -helical motifs in the assembled structure (Fig. 1b–d), including the CTD dimerization helix, H9, with a distinct kink in the middle (Fig. 1d). The linker region (residues 146–150), followed by a  $3_{10}$ -helix, shows strong electron density (Fig. 1d), suggesting that it adopts a relatively ordered and stable conformation. This linker region is a critical determinant for proper capsid assembly and stability<sup>8</sup>. The map also shows detailed intermolecular interactions at all four interfaces critical for assembly: the inter-hexamer CTD interfaces at the pseudo-two-fold and pseudo-three-fold axes, and the hexamer forming NTD-NTD and NTD-CTD interfaces (Fig. 1c and Supplementary Fig. 1c)<sup>9,11,14</sup>.

The 8 Å resolution density map allowed reliable atomic modelling for the entire assembly, using molecular dynamics flexible fitting (MDFF)<sup>15</sup>. We docked atomic structures (3H47 for CA NTD<sup>9</sup> and 2KOD for CA CTD<sup>11</sup>) into the electron density, built the linker and missing loops using homology modelling, and then applied MDFF



**Figure 1 | Cryo-EM reconstruction of HIV-1 CA tubular assembly at 8 Å resolution and MDFF.** **a**, A cryo-EM image of recombinant A92E CA tubular assembly. Scale bar, 100 nm. **b**, Electron density map of the A92E CA tube with  $(-12, 11)$  helical symmetry. Yellow arrows indicate pairs of helix H9, located between adjacent hexamers. **c**, MDFF model of the HIV-1 capsid assembly, superimposed with the electron density map contoured at  $4.0\sigma$ . Three CA hexamers, with NTDs (blue) and CTDs (orange), are shown. **d**, MDFF model of a CA monomer viewed from two angles. **e**, Two CTD dimer structures along  $-1$  (orange) and  $11$  (yellow) helical directions, superimposed onto the NMR solution dimer structure (grey, 2KOD).

<sup>1</sup>Department of Structural Biology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15260, USA. <sup>2</sup>Pittsburgh Center for HIV Protein Interactions, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15260, USA. <sup>3</sup>Department of Physics and Beckman Institute, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA. <sup>4</sup>Department of Pathology, Microbiology and Immunology, Vanderbilt University School of Medicine, Nashville, Tennessee 37232, USA. <sup>5</sup>Department of Physics, University of Central Florida, Orlando, Florida 32816, USA.

<sup>6</sup>Department of Mechanical Engineering and Materials Science, Swanson School of Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA.

\*These authors contributed equally to this work.



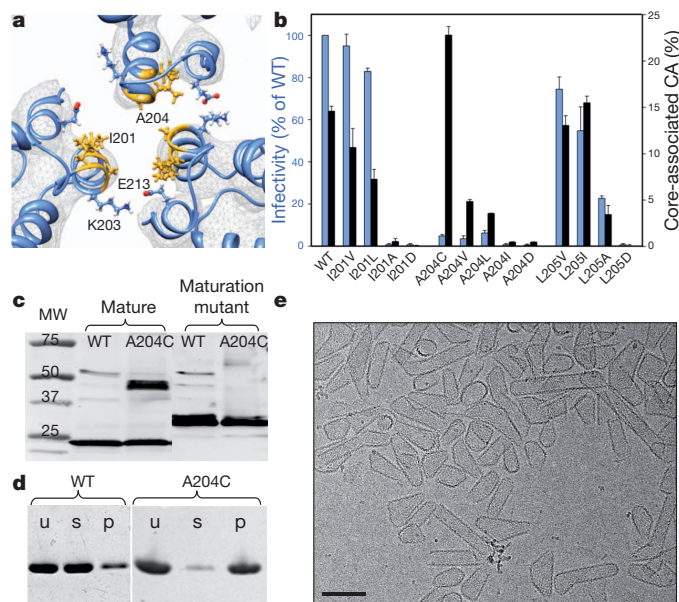
(Supplementary Video 1). The resulting atomic model for the tubular assembly comprises 71 CA hexamers and 13 million atoms, including water molecules and ions. The final model fits the experimental map remarkably well, as indicated by the excellent overlap between  $\alpha$ -helices and their corresponding rod-like densities (Fig. 1c, d), with an overall cross-correlation function value of 0.96. All the densities are accounted for in the model, including those of the linker region (Fig. 1d) and  $\beta$ -hairpin region. An equivalent pseudo-atomic model was obtained independently, using a normal mode flexible-fitting method (Supplementary Fig. 2a).

The MDFF model of the full CA assembly contains an almost invariant hexameric NTD structure with an average root mean squared deviation of  $1.5 \pm 1.0$  Å for backbone atoms between all 71 hexamers. The hexameric arrangement of the NTD is very similar to that observed in the crystal structure of cross-linked hexameric CA protein<sup>9</sup> (Supplementary Fig. 2b). Further, the intermolecular interactions in a single hexameric unit are the same as those observed in the crystal structure (Supplementary Fig. 2c, d). Considering that retroviral CA proteins share a common NTD fold and all form hexamers, hexameric CA NTD appears to constitute a building block in the final assembly.

In contrast, the CTD dimers, which connect adjacent assembly units on the inner, more curved surface, are more variable than the NTD interfaces, with a root mean squared deviation of  $3.0 \text{ Å} \pm 0.9 \text{ Å}$  for all 71 dimers. Three distinct CTD dimer conformations are seen in the assembled helical tube, with dimerization helix H9 crossing angles of  $36^\circ$ ,  $44^\circ$  or  $54^\circ$ . Several conformers for the CTD dimer and full-length dimeric CA were previously observed in solution<sup>16</sup>, demonstrating the plastic nature of the dimer interface. Nonetheless, key contacts are formed by the hydrophobic residues, W184 and M185, albeit with small variations in their local environment (Fig. 1e). The CTD girdle of the CA hexamer, helices H9 and H10 in particular, seems more

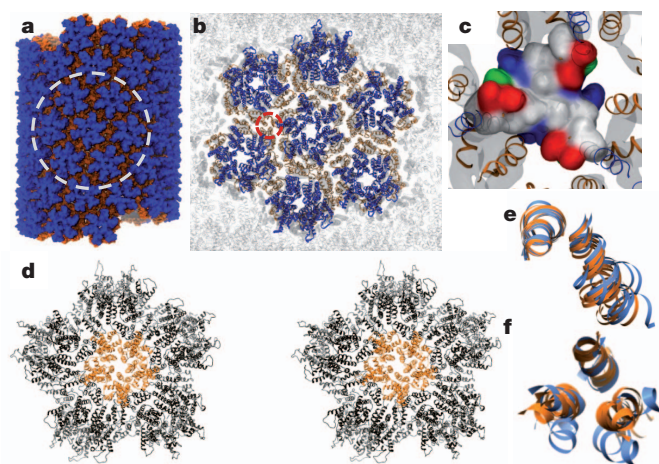
variable, compared with the CTD seen in the crystal structure (Supplementary Fig. 2b, e)<sup>9</sup>. Thus, the curvature needed for the helical tube assembly is accommodated by variable orientations in the CTD, mediated by the intrinsic flexibility of helix H9 in the monomer.

Important new structural detail is seen at the CTD trimer interface. The MDFF model of the tubular assembly shows that this interface involves a patch of hydrophobic residues, specifically I201, L202, A204 and L205, all situated on one face of the amphipathic CTD helix H10 (Supplementary Fig. 3a, b). At the centre of the hydrophobic core reside residues I201 and A204 (Fig. 2a and Supplementary Fig. 3b). These residues contribute a large portion of the total buried surface area ( $8,758 \text{ Å}^2$ ) at the pseudo-three-fold axis. The central hydrophobic core is surrounded by polar residues of opposite charges, for example K203 and E213 (Figs 2a and 3c), further stabilizing the interface. Residues K203 and E213 are highly conserved and were previously shown to be critical for optimal HIV-1 capsid stability and for viral infectivity<sup>11,17</sup>. To test further the involvement of the hydrophobic interactions in HIV-1 capsid function, we performed mutagenesis and chemical crosslinking studies. Remarkably, hydrophobic replacements (I201V/L, A204V/L and L205V/I) largely retained the infectivity and capsid stability (Fig. 2b)<sup>18</sup>, but mutant virions with polar residue substitutions (I201D, A204D and L205D) were non-infectious (Fig. 2b) and contained unstable or abnormal cores (Supplementary Fig. 3c–g). Interestingly, these mutations had no apparent effect on the morphology of immature virions (Supplementary Fig. 3h–j). Purified CA proteins with the same polar substitutions failed to assemble *in vitro* (Supplementary Fig. 3m, n). The A204C CA mutant, which exhibits a Cys-Cys backbone distance of  $5.3 \text{ Å}$  between adjacent molecules at the pseudo-three-fold axis, exhibited disulphide crosslinking of CA into dimers, *in vitro* and in mature virions (Fig. 2c). This was not the case for maturation-defective virions in which cleavage of nucleocapsid protein was prohibited (Fig. 2c), implying that the hydrophobic CTD trimer interface is specific to the mature capsid<sup>19</sup>. In addition, the A204C CA protein exhibited higher assembly efficiency ( $>90\%$ ) compared with wild-type CA (10–15%) (Fig. 2d) and assembled under physiological salt concentrations. More intriguingly, the A204C protein (and A204L) assembled into closed cones and short tubes that closely resembled authentic lentiviral cores (Fig. 2e and Supplementary Fig. 3l, o), rather than the long tubes typically formed by wild-type



**Figure 2 | Mutational analysis of the hydrophobic trimer interface.**

**a**, Detailed view of the trimer interface, with the structural model superimposed on the density map (grey mesh, contoured at  $3.5\sigma$ ). Selected residues are depicted in stick-and-ball representation. **b**, Virus infectivity (blue) and capsid stability (black), as percentage of total CA,  $\pm$  standard deviation ( $n = 3$ ), of wild-type (WT) and CA trimer interface mutants. **c**, Spontaneous disulphide crosslinking of A204C mature and maturation-defective virions analysed by immunoblotting for CA. **d**, *In vitro* assembly of recombinant wild-type and A204C CA proteins, assayed with high-speed sedimentation and polyacrylamide gel electrophoresis analysis. Letters u, s and p denote the unassembled reaction mixture (u) and the supernatant (s) and pellet (p) after assembly. **e**, Cryo-EM image of A204C *in vitro* assembly. Scale bar, 100 nm.



**Figure 3 | All-atom molecular dynamics simulation of CA tubular assembly.**

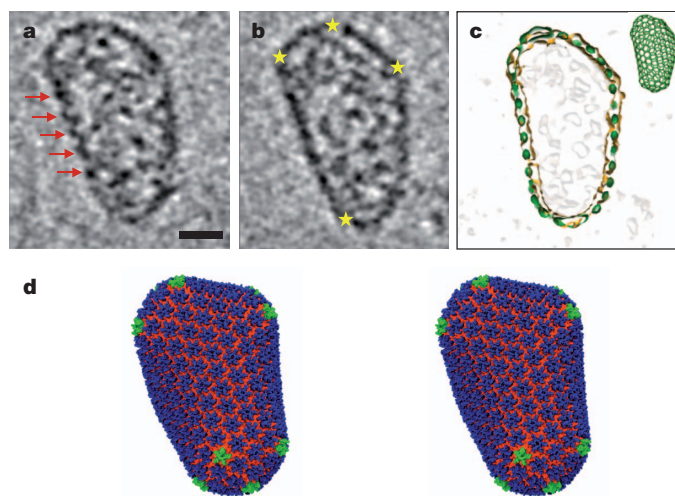
**a**, All-atom tubular assembly model comprising 71 CA hexamers (blue NTD and orange CTD) equilibrated for 125 ns. **b**, Ribbon representation of the tubular assembly, highlighting an HOH, circled area in **a**, superimposed on the density map (grey). **c**, Surface representation of the trimer interface (circled area in **b**) in a 425 ns equilibrated HOH model (hydrophobic, polar, negative and positive residues are grey, green, red and blue, respectively). **d**, Stereo view of a POH model after  $1.5 \mu\text{s}$  equilibration. A pentamer (orange) is surrounded by five hexamers (black). **e**, **f**, Superposition of CTD H9 dimer (**e**) or H10 trimer (**f**) interfaces from HOH (blue) and POH (orange).



CA (Supplementary Fig. 3k). This suggests that CA pentamers are readily incorporated in A204C assemblies. Consistent with the *in vitro* results, virions carrying the A204C substitution showed enhanced capsid stability and reduced infectivity (Fig. 2b). These results indicate that hydrophobic interactions at the CTD trimer interface play a critical role in mature capsid assembly and stability. This role is further supported by semi-three-dimensional Monte Carlo simulations of 36 CA dimers, on the basis of the assembly phase diagram established previously (Supplementary Fig. 4)<sup>20</sup>.

A 'fullerene cone' model of the HIV-1 capsid requires insertion of 12 CA pentamers into the hexagonal surface lattice to close the ovoid. Although the atomic structures of pentameric and hexameric HIV-1 CA mutants have been determined<sup>9,10</sup>, and their structural similarity suggests that they are quasi-equivalent<sup>10</sup>, detailed contacts between pentamers and hexamers for building a closed fullerene cone are not known. On the basis of the hexamer-of-hexamers (HOH) motif (six hexamers surrounding a central hexamer) extracted from the helical tube assembly (Fig. 3a, b), we generated a model for a pentamer of hexamers (POH; five hexamers encompassing a central pentamer), by replacing the central hexamer with a pentamer. The resulting model contains gaps between the surrounding hexamers because of the deletion of one hexamer, and poor dimer and trimer interactions (Supplementary Video 2). However, during a molecular dynamics simulation (Supplementary Table 1), these gaps closed quickly (Supplementary Video 2), converting the relatively flat starting model into a highly curved dome-like structure (Fig. 3d, Supplementary Fig. 5a, b and Supplementary Video 3). More specifically, sharp bite angles between adjacent subunits were formed, namely 137° and 147° for pentamer-hexamer and hexamer-hexamer edges, respectively, compared with 167° and 147° in the case of HOH (Supplementary Fig. 5b). The sharp bite angle closely matches the 135° value seen in the fullerene cone geometric model<sup>21</sup>. All pre-existing intra-hexamer (or pentamer) and CTD dimer intermolecular interfaces were preserved in the simulation, whereas new dimer and trimer interactions between hexamers and between pentamer and hexamers were formed in 300 ns (Fig. 3d and Supplementary Video 2), validating the previously suggested quasi-equivalence in the capsid lattice<sup>22,23</sup>. Although the CTD dimer interface remained slightly variable, as in the HOH model (Fig. 3e), the three-helix bundle at the CTD trimer interface was conserved but more closely packed in the POH (Fig. 3f and Supplementary Fig. 5c, orange) compared with the HOH (Fig. 3f, blue). This is intriguing and suggests that enhancing the interaction at the three-fold axis, for example by A204C crosslinking, facilitates incorporation of pentamers into the assembly (Fig. 2e). Electrostatic destabilization at the six-fold axis in CA NTD<sup>10,22</sup>, counterbalanced by the stabilizing hydrophobic interactions at the three-fold axis, probably controls pentamer incorporation. Given that the trimer contact is absent in the planar two-dimensional crystal of the CA assembly<sup>14</sup>, but clearly present in the helical tubular assembly, and even tighter in the highly curved POH model, we suggest that the three-fold interface is essential for the curved asymmetric assembly of the conical capsid.

The HOH and POH assembly structures allow construction of an improved geometric HIV-1 capsid model<sup>10</sup>. Building a realistic atomic model of the HIV-1 capsid, however, requires an experimentally determined native HIV-1 core structure. Because of the non-uniform, asymmetric nature of the native cores, we performed structural analysis by cryo-electron tomography (cryo-ET) of HIV-1 cores isolated from virions (Supplementary Fig. 6 and Supplementary Video 4). The tomographic slices of a best-quality HIV-1 core, with a cone angle of 23° and representing the most common core<sup>24</sup> (Supplementary Fig. 6b), clearly exhibit arrays of CA hexamers (Fig. 4a, red arrows) and declinations made by CA pentamers (Fig. 4b, yellow stars) at the outer capsid density layer. The native core also includes internal densities corresponding to the viral RNA genome and proteins enclosed within the capsid (Fig. 4a, b). At the current tomography resolution, individual hexamers or pentamers in the capsid cannot be discerned.



**Figure 4 | All-atom HIV-1 capsid model.** **a, b**, Cryo-ET analysis of an isolated, native HIV-1 core, shown as two representative slices through the three-dimensional volume. Red arrows indicate arrays of CA hexamers; yellow stars indicate locations of sharp curvature change. Scale bar, 20 nm. **c**, A fullerene cone model (216H + 12P, green inset) matches the shape and size of the capsid, shown by the overlay of densities from the segmented capsid (orange contour) and the fullerene model (green). HIV-1 core internal densities are shown in light grey. **d**, Stereo view of the final molecular dynamics equilibrated all-atom capsid model (model I, see text) comprising 216 CA hexamers (blue, NTD; orange, CTD) and 12 CA pentamers (green).

Thus, guided by the shape, size and structural features of the capsid layer, extensive fullerene models, including classes containing 252, 216, 186 or 166 hexamers, were evaluated and top candidates from each class (119 of a total of 724) were tested by cross-correlation between the model and the capsid density (Supplementary Fig. 7). Two fullerene models, I and II, comprising 12 pentamers, as well as 216 and 186 hexamers, respectively, were selected to generate all-atom molecular dynamics models of the entire HIV-1 capsid, comprising 1,356 (model I, Fig. 4c, d) or 1,176 (model II, Supplementary Fig. 7b) CA subunits. These models were confirmed and analysed through a fully solvated, unconstrained 100 ns, 64 million atom, molecular dynamics simulation as described in Supplementary Material (Supplementary Table 1).

The resulting capsid model I contains a variable CTD structure, with an average root mean squared deviation of  $3.6 \pm 0.5$  Å for backbone atoms between all 216 hexamers, compared with  $2.3 \pm 0.4$  Å for the NTD. Analysis of the bite angle between the neighbouring hexamers (or pentamers) shows a bimodal distribution (Supplementary Fig. 8a), with a minor population of sharp bite angles (mean = 139°) located around pentamers (Supplementary Fig. 8b), in addition to most continuously varying bite angles (144–180°). Compared with the recent cryo-EM structure of an immature intermediate of a different retrovirus<sup>25</sup>, our mature capsid structure shows a very different set of inter-subunit interactions, suggesting large conformational changes accompany retrovirus maturation. Our atomic models of the complete HIV-1 capsid further highlight the three-fold CA CTD trimer interface as an attractive therapeutic target and provide a platform for future studies of capsid function and for targeted pharmacological intervention.

## METHODS SUMMARY

CA tubes were assembled at 2 mg ml<sup>-1</sup> in a buffer containing 1 M NaCl and 50 mM Tris-HCl, pH 8.0. Cryo-EM micrographs were collected on Kodak SO-163 films, on an FEI Polara microscope operated at 200 kV, and digitized at a pixel size of 1.09 Å. Iterative Helical Real Space Reconstruction<sup>13</sup> was used to estimate the helical parameters. The final refinement was performed using a previously established method<sup>12</sup>. Cryo-ET tilt series of native HIV-1 cores were collected on a Gatan 4k × 4k CCD (charge-coupled device) camera by tilting the specimen from -70° to 66°, with a total dose of approximately 120 e<sup>-</sup>/Å<sup>2</sup> and a defocus value around

8  $\mu\text{m}$ . Three-dimensional tomograms were reconstructed using IMOD and TOMO3D software<sup>26,27</sup>. MDFF<sup>15</sup> was applied for 10 ns using NAMD2.9 (ref. 28) at 2 fs time-steps and with helical symmetry restraints. The MDFF-derived HOH structure was further equilibrated for 425 ns. On the basis of the equilibrated HOH model and a published pentamer structure 3P05 (ref. 10), a POH model was built and equilibrated for 1.5  $\mu\text{s}$ . A total of 71 MDFF-derived hexamers were docked into the cryo-EM map and equilibrated for 125 ns. Simulations of the complete HIV-1 capsid (64 million atoms, 100 ns) were performed with NAMD. Virus particles were derived by transfection of the full-length HIV-1 proviral construct R9 and mutant derivatives into 293T cells, and infectivity was quantified by titration on HeLa-P4 cells. Capsid stability was determined by measuring the amount of CA protein present after purification of HIV-1 cores, and the rate of uncoating was analysed by quantifying CA release from purified cores at 37 °C. For crosslinking analysis, particles were pelleted from the supernatants of transfected 293T cells, analysed by non-reducing SDS-PAGE, and CA was detected by immunoblotting.

Received 2 November 2012; accepted 5 April 2013.

- Sundquist, W. I. & Krausslich, H. G. HIV-1 assembly, budding, and maturation. *Cold Spring Harb. Perspect. Med.* **2**, a006924 (2012).
- Ganser, B. K., Li, S., Klishko, V. Y., Finch, J. T. & Sundquist, W. I. Assembly and analysis of conical models for the HIV-1 core. *Science* **283**, 80–83 (1999).
- Li, S., Hill, C. P., Sundquist, W. I. & Finch, J. T. Image reconstructions of helical assemblies of the HIV-1 CA protein. *Nature* **407**, 409–413 (2000).
- Gitti, R. K. *et al.* Structure of the amino-terminal core domain of the HIV-1 capsid protein. *Science* **273**, 231–235 (1996).
- Momany, C. *et al.* Crystal structure of dimeric HIV-1 capsid protein. *Nature Struct. Biol.* **3**, 763–770 (1996).
- Du, S. *et al.* Structure of the HIV-1 full-length capsid protein in a conformationally trapped unassembled state induced by small-molecule binding. *J. Mol. Biol.* **406**, 371–386 (2011).
- Gamble, T. R. *et al.* Structure of the carboxyl-terminal dimerization domain of the HIV-1 capsid protein. *Science* **278**, 849–853 (1997).
- Jiang, J. *et al.* The interdomain linker region of HIV-1 capsid protein is a critical determinant of proper core assembly and stability. *Virology* **421**, 253–265 (2011).
- Pornillos, O. *et al.* X-ray structures of the hexameric building block of the HIV capsid. *Cell* **137**, 1282–1292 (2009).
- Pornillos, O., Ganser-Pornillos, B. K. & Yeager, M. Atomic-level modelling of the HIV capsid. *Nature* **469**, 424–427 (2011).
- Byeon, I. J. *et al.* Structural convergence between cryo-EM and NMR reveals intersubunit interactions critical for HIV-1 capsid function. *Cell* **139**, 780–790 (2009).
- Meng, X., Zhao, G. & Zhang, P. Structure of HIV-1 capsid assemblies by cryo-electron microscopy and iterative helical real-space reconstruction. *J. Vis. Exp.* **54**, e3041 (2011).
- Egelman, E. H. The iterative helical real space reconstruction method: surmounting the problems posed by real polymers. *J. Struct. Biol.* **157**, 83–94 (2007).
- Ganser-Pornillos, B. K., Cheng, A. & Yeager, M. Structure of full-length HIV-1 CA: a model for the mature capsid lattice. *Cell* **131**, 70–79 (2007).
- Trabuco, L. G., Villa, E., Schreiner, E., Harrison, C. B. & Schulten, K. Molecular dynamics flexible fitting: a practical guide to combine cryo-electron microscopy and X-ray crystallography. *Methods* **49**, 174–180 (2009).
- Byeon, I. J. *et al.* Motions on the millisecond time scale and multiple conformations of HIV-1 capsid protein: implications for structural polymorphism of CA assemblies. *J. Am. Chem. Soc.* **134**, 6455–6466 (2012).
- Forshey, B. M., von Schwedler, U., Sundquist, W. I. & Aiken, C. Formation of a human immunodeficiency virus type 1 core of optimal stability is crucial for viral replication. *J. Virol.* **76**, 5667–5677 (2002).
- Joshi, A., Nagashima, K. & Freed, E. O. Mutation of dileucine-like motifs in the human immunodeficiency virus type 1 capsid disrupts virus assembly, gag-gag interactions, gag-membrane binding, and virion maturation. *J. Virol.* **80**, 7939–7951 (2006).
- Meng, X. *et al.* Protease cleavage leads to formation of mature trimer interface in HIV-1 capsid. *PLoS Pathog.* **8**, e1002886 (2012).
- Chen, B. & Tycko, R. Simulated self-assembly of the HIV-1 capsid: protein shape and native contacts are sufficient for two-dimensional lattice formation. *Biophys. J.* **100**, 3035–3044 (2011).
- Yeager, M. Design of *in vitro* symmetric complexes and analysis by hybrid methods reveal mechanisms of HIV capsid assembly. *J. Mol. Biol.* **410**, 534–552 (2011).
- Cardone, G., Purdy, J. G., Cheng, N., Craven, R. C. & Steven, A. C. Visualization of a missing link in retrovirus capsid assembly. *Nature* **457**, 694–698 (2009).
- Caspar, D. L. & Klug, A. Physical principles in the construction of regular viruses. *Cold Spring Harb. Symp. Quant. Biol.* **27**, 1–24 (1962).
- Briggs, J. A., Wilk, T., Welker, R., Krausslich, H. G. & Fuller, S. D. Structural organization of authentic, mature HIV-1 virions and cores. *EMBO J.* **22**, 1707–1715 (2003).
- Bharat, T. A. *et al.* Structure of the immature retroviral capsid at 8 Å resolution by cryo-electron microscopy. *Nature* **487**, 385–389 (2012).
- Mastrorade, D. N. Correction for non-perpendicularity of beam and tilt axis in tomographic reconstructions with the IMOD package. *J. Microsc.* **230**, 212–217 (2008).
- Agulleiro, J. I. & Fernandez, J. J. Fast tomographic reconstruction on multicore computers. *Bioinformatics* **27**, 582–583 (2011).
- Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank P. Schwerdtfeger for access to his software for creating fullerene models, T. Brosenitsch for reading the manuscript, and M. DeLucia and J. Mehrens for technical assistance. This work was supported by the National Institutes of Health (GM082251, GM085043 and GM104601) and the National Science Foundation (PHY0822613, MCB0744057). Large-scale molecular dynamics simulations were performed on the Blue Waters Computer, financed by the National Science Foundation (OCI 07-25070).

**Author Contributions** G.Z., J.R.P., E.L.Y., A.M.G., K.S., C.A. and P.Z. designed the research. J.N. and J.A. prepared samples for electron microscopy. G.Z. collected cryo-EM data. G.Z., X.M. and P.Z. analysed cryo-EM and cryo-ET data. E.L.Y. and C.A. performed biochemical and functional analysis. K.S. developed large-scale modelling methodology; J.R.P. performed molecular dynamics simulations and B.C. performed CG-MC simulations. G.Z., J.R.P., K.S. and P.Z. analysed atomic models. G.Z., J.R.P., A.M.G., K.S., C.A. and P.Z. wrote the paper with support from all the authors.

**Author Information** Cryo-EM structural data have been deposited in the EMDDataBank under accession codes EMD-5582 and EMD-5639, and the MDFF atomic model of the CA HOH and models of HIV-1 capsid have been deposited in the Protein Data Bank under accession numbers 3J34, 3J3Q, 3J3Y. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.Z. (pez7@pitt.edu), C.A. (chris.aiken@vanderbilt.edu) or K.S. (kschulte@ks.uiuc.edu).

# Crystal structure of a nitrate/nitrite exchanger

Hongjin Zheng<sup>1</sup>, Goragot Wisedchaisri<sup>1</sup> & Tamir Gonen<sup>1</sup>

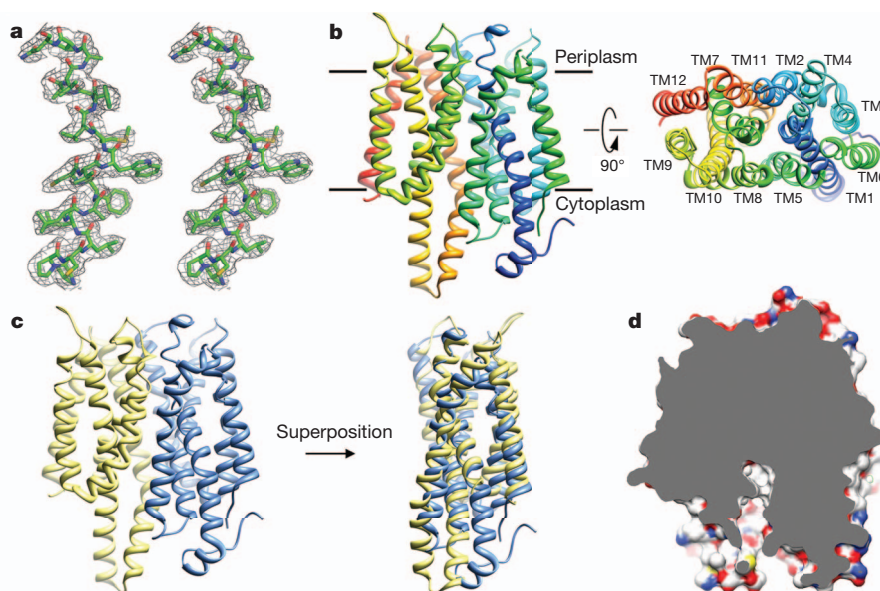
Mineral nitrogen in nature is often found in the form of nitrate ( $\text{NO}_3^-$ ). Numerous microorganisms evolved to assimilate nitrate and use it as a major source of mineral nitrogen uptake<sup>1</sup>. Nitrate, which is central in nitrogen metabolism, is first reduced to nitrite ( $\text{NO}_2^-$ ) through a two-electron reduction reaction<sup>2,3</sup>. The accumulation of cellular nitrite can be harmful because nitrite can be reduced to the cytotoxic nitric oxide. Instead, nitrite is rapidly removed from the cell by channels and transporters, or reduced to ammonium or dinitrogen through the action of assimilatory enzymes<sup>3</sup>. Despite decades of effort no structure is currently available for any nitrate transport protein and the mechanism by which nitrate is transported remains largely unknown. Here we report the structure of a bacterial nitrate/nitrite transport protein, NarK, from *Escherichia coli*, with and without substrate. The structures reveal a positively charged substrate-translocation pathway lacking protonatable residues, suggesting that NarK functions as a nitrate/nitrite exchanger and that protons are unlikely to be co-transported. Conserved arginine residues comprise the substrate-binding pocket, which is formed by association of helices from the two halves of NarK. Key residues that are important for substrate recognition and transport are identified and related to extensive mutagenesis and functional studies. We propose that NarK exchanges nitrate for nitrite by a rocker switch mechanism facilitated by inter-domain hydrogen bond networks.

The nitrate/nitrite porter (NNP) family of membrane proteins evolved to efficiently translocate the ionic molecules  $\text{NO}_3^-$  and  $\text{NO}_2^-$  across the membrane<sup>4,5</sup>. Two nitrate/nitrite transport proteins NarK and

NarU were identified in *Escherichia coli*<sup>6–10</sup>. NarK proteins have been shown to catalyse either nitrate/nitrite exchange or nitrate uptake, presumably by symport with a proton<sup>1,9,10</sup>. The former activity would be associated with respiration, whereas the latter could be associated either with respiration or the net assimilation of nitrite into cell material. (An additional membrane protein, NirC, functions as a  $\text{H}^+$ /nitrite channel in *E. coli* but is not a member of the NNP<sup>10</sup>.) Nitrate is the preferred source of nitrogen for plants and at least 16 nitrate/nitrite transport proteins have been identified<sup>11</sup>. In plants the function of NNP proteins is probably related solely to the net assimilation of nitrogen.

The NNP family belongs to the major facilitator superfamily (MFS) of secondary transporters. MFS members show specificity to a wide range of molecules<sup>12</sup>. Although more than 58 distinct families of transporters make up the MFS, representatives of only six such families have been crystallized and their structure determined<sup>13–19</sup>. These six representative transporters require protons for their function. All MFS members are postulated to function through the rocker switch mechanism<sup>12</sup>. They all share a common structural topology, but share little or no sequence homology.

Here we report the first crystal structure of a nitrate transport protein that also transports nitrite. The structure of the *E. coli* NarK with and without substrate was determined by X-ray crystallography. Functionally important residues that form the substrate-binding pocket are identified and related to previously described mutagenesis and functional studies. We provide the first evidence that NarK functions as a nitrate/nitrite exchanger, and that protons are probably not co-transported with the substrate.

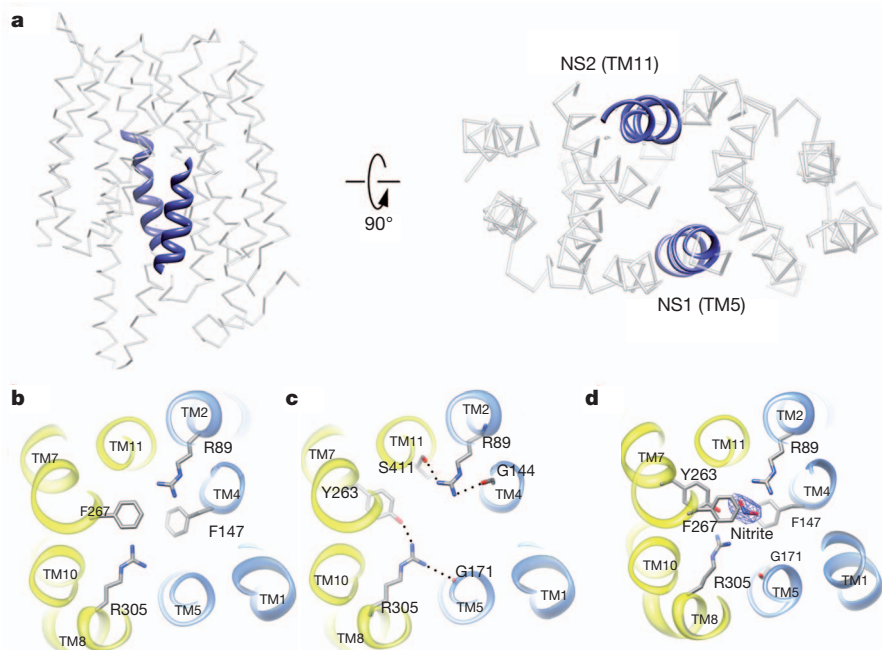


**Figure 1 | The crystal structure of NarK.** **a**, Part of TM2 of NarK is shown in stereo view with a sigmaA-weighted  $2F_o - F_c$  map at 2.6 Å resolution, contoured at  $1.0\sigma$ . **b**, Left, NarK structure viewed from the plane of membrane with the putative location of the lipid bilayer as indicated. NarK is coloured in rainbow with the N terminus in blue. Right, NarK viewed from the

periplasmic side. The identity of the 12 transmembrane helices is indicated. **c**, The N-terminal domain (TM1–TM6) of NarK (blue) is pseudo-symmetric to the C-terminal domain (TM7–TM12) (yellow) and can be superimposed with an r.m.s.d. of 2.9 Å. **d**, Cut-away surface representation of the inward-facing NarK shows the central cavity exposed to the cytosol.

<sup>1</sup>Janelia Farm Research Campus, Howard Hughes Medical Institute, 19700 Helix Drive, Ashburn, Virginia 20147, USA.





**Figure 2 | The substrate-binding site in NarK.** **a**, Two highly conserved nitrate signature motifs in TM5 and TM11 (blue helices) at the centre of NarK form the nitrate/nitrite transport pathway. **b**, The substrate-binding pocket is defined by two evolutionarily conserved and functionally important arginine residues R89 and R305. The binding site is capped above and below by F267 and F147, respectively. **c**, R89 and R305 are stabilized by inter-domain hydrogen bonds as depicted. The two halves of NarK are indicated as blue

We overexpressed NarK, and purified the protein to homogeneity as described in the Methods section. Well-ordered high-quality crystals were obtained when NarK was co-crystallized with the Fab fragment of a monoclonal antibody we developed. The data extended to 2.6 Å were phased using molecular replacement with Fab as a search model. The resulting density map was of high enough quality, allowing us to build and refine the NarK structure (Fig. 1a). The asymmetric unit contains one molecule of NarK forming a complex with one molecule of Fab (Supplementary Fig. 1). NarK, as other MFS proteins, is structurally divided into two domains, the amino-terminal half and the carboxy-terminal half each consisting of six transmembrane helices (TM1–TM6 and TM7–TM12, respectively) (Fig. 1b, c). The two domains are connected by a long loop between TM6 and TM7 (disordered in our structure), and it is thought that the substrate transport pathway is localized at the interface between these two domains. NarK appears to be in the inward-facing conformation as the hydrophilic central cavity is exposed to the cytosolic side (Fig. 1d). A detailed description of the crystal packing and the overall architecture of NarK can be found in the Supplementary Information.

All members of the NNP family contain two stretches of conserved residues called the nitrate signature (NS) motifs (Supplementary Fig. 2). The nitrate signature motifs are not found in other MFS members, but are a unique feature of the NNP family<sup>20</sup>. In NarK, the NS1 motif is formed by residues 164–175 (GGALGLNGGLGN) located on TM5. The NS2 motif of NarK is formed by residues 408–420 (GFISAIGAIGGFF) located on TM11 (Fig. 2a, blue). The nitrate signature motifs in NarK are located at the centre of the protein, lining part of the substrate transport pathway (Fig. 2a, right). Both of the nitrate signature motifs are glycine-rich, which ensures a tight fit among the surrounding helices. The result is a significantly more compact structure for NarK when compared to other known structures of MFS members (Supplementary Fig. 3).

To transport anions like  $\text{NO}_3^-$  and  $\text{NO}_2^-$ , polar residues lining the central pore are most likely to form the substrate-binding pocket and be

involved in substrate recognition and transport. Two arginine residues, R89 from TM2 and R305 from TM8, are absolutely conserved among all nitrate/nitrite transporters in both prokaryotes and eukaryotes (Supplementary Fig. 2). Structurally, R89 and R305 are in plane and appose one another at the very centre of NarK, with their side chains extending well into the central cavity of the transporter (Fig. 2b). These arginines are capped by two phenylalanine residues: F267 above and F147 below. Together the arginines and the phenylalanines form the substrate-binding pocket. The only bulky side chain in plane with the arginines is Y263, which forms a hydrogen bond with R305 (Fig. 2c). The two arginine side chains are stabilized by an intricate system of inter-domain hydrogen bonds that link the two halves of NarK (Fig. 2c). Mutation of the residues described above led to a complete loss of function in NarK and its close homologues (Supplementary Fig. 2

**Table 1 | Mutagenesis and functional study of key residues important for nitrate/nitrite exchange**

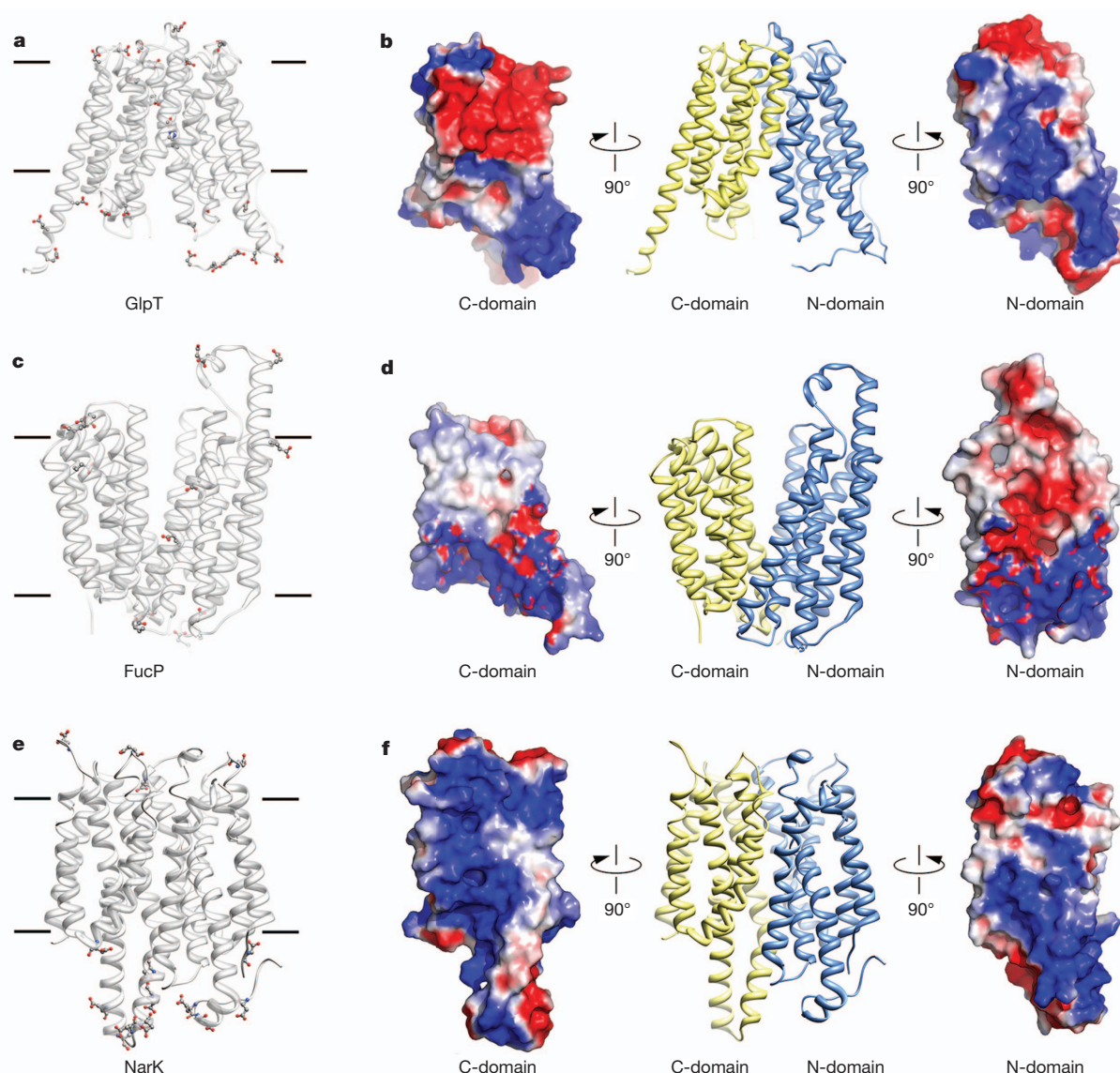
NarK ( <i>E. coli</i> )	NarU ( <i>E. coli</i> )	NrTA ( <i>Aspergillus nidulans</i> )	Colony growth on nitrate
<b>R89</b>	R87K* R87P*	R87K† R87T†	+
<b>N175</b>		N168A‡ N168Q† N168C†	–
<b>Y263</b>	Y261Q* Y261N*		+
<b>R305</b>	R303Q* R303C*	R368K† R368Q† R368C†	+
<b>A415</b>		N459A‡ N459Q† N459K† N459C†	–

\*Study published in ref. 10.

†Study published in ref. 24.

‡Study published in ref. 21.

See references above for a comprehensive list of all available mutagenesis data.



**Figure 3 | Protons are probably excluded from the substrate translocation pathway of NarK.** **a**, Location of histidine, aspartate and glutamate residues in the anion transporter GlpT. Acidic residues line the substrate translocation pathway. **b**, Electrostatic surface representation for each domain of GlpT showing a relatively even distribution of positive and negative charges in the substrate translocation pathway. **c**, Location of histidine, aspartate and glutamate residues in the fucose transporter FucP. Acidic residues line the

substrate translocation pathway. **d**, Electrostatic surface representation for each domain of FucP. **e**, Location of histidine, aspartate and glutamate residues in NarK. No acidic residues are found in the substrate translocation pathway of NarK. **f**, Electrostatic surface representation for NarK showing a dominantly positively charged substrate translocation pathway. It represents a formidable barrier for the translocation of protons but could attract negatively charged molecules like nitrate and nitrite.

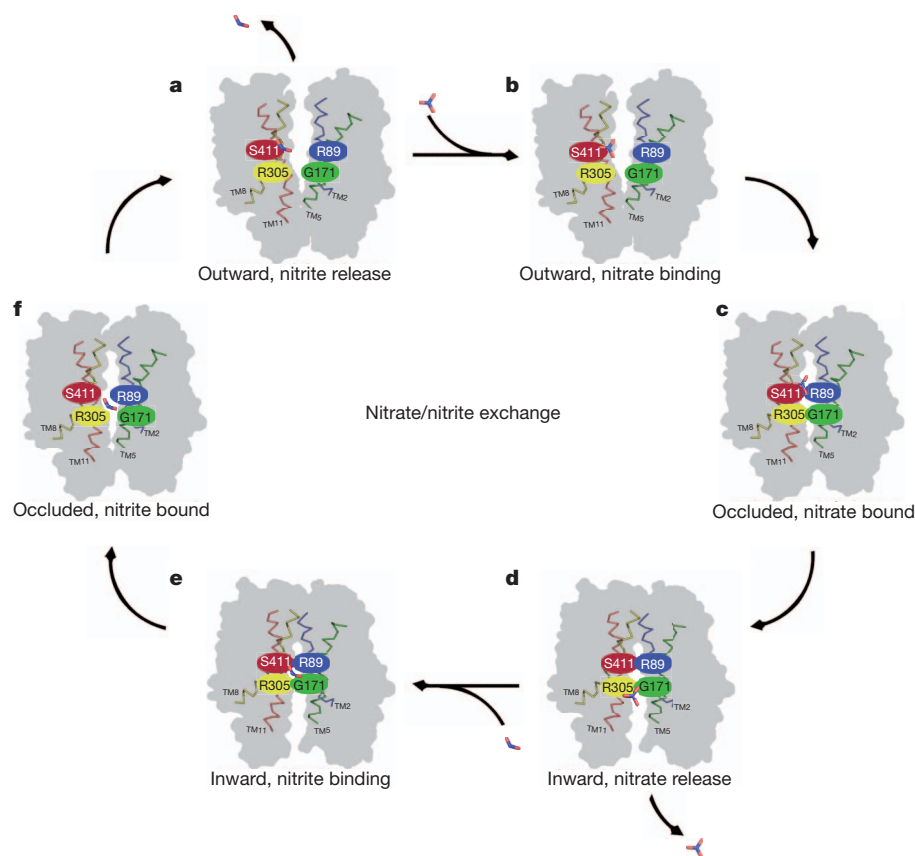
and Table 1). A detailed discussion of the relevant mutational and functional studies can be found in Supplementary Information.

Soaking our NarK crystals with sodium nitrate deteriorated the crystal packing and did not yield meaningful data. In sharp contrast, soaking the crystals with sodium nitrite did not significantly affect crystallinity and yielded data to 2.8 Å resolution (Supplementary Table 1), allowing us to visualize the nitrite bound in the substrate-binding pocket (Fig. 2d and Supplementary Fig. 4). Overall the structure of substrate-free NarK and nitrite-bound NarK are very similar, having an all C $\alpha$ -atom root mean squared deviation (r.m.s.d.) of 0.6 Å (Supplementary Fig. 5). This is not surprising because NarK is probably stabilized in the inward-facing conformation by crystal contacts and the Fab could further restrict protein movement (Supplementary Fig. 1). Nevertheless, clear densities were observed for nitrite in the substrate-binding pocket (Fig. 2d and Supplementary Fig. 4). Nitrite was observed in-plane with R89 and R305 at the substrate-binding pocket where it is capped above

and below by F267 and F147, respectively (Fig. 2d). This binding configuration stabilizes the substrate via the  $\pi$ -electron delocalization among the arginine and phenylalanine side chains. Arginine R305 changed its conformation so that the inter-domain hydrogen bond network involving Y263 and G171 was disrupted upon nitrite binding.

Although it is clear that NarK and other NNP members are capable of  $\text{NO}_3^-$  uptake and  $\text{NO}_2^-$  export, it is not clear what the mechanism is or whether the process is proton coupled. Three distinct modes of action have been proposed:  $\text{H}^+/\text{NO}_3^-$  symport,  $\text{H}^+/\text{NO}_2^-$  antiport, or a  $\text{NO}_3^-/\text{NO}_2^-$  exchange without  $\text{H}^+$  translocation<sup>1,9,10</sup>.

Typically, channels and transporters that translocate protons use residues that are capable of protonation or deprotonation along the pore or substrate pathway<sup>12</sup>. For example, the lactose permease LacY, which co-transporters lactose with protons, uses glutamate and histidine residues to translocate the protons<sup>21</sup>. The glycerol-3-phosphate/phosphate antiporter GlpT uses a protonated histidine to facilitate substrate



**Figure 4 | Proposed mechanism of nitrate/nitrite exchange.** a–f, Six conformations of NarK are depicted as outward facing (a), outward facing with nitrate bound (b), occluded with nitrate bound (c), inward facing with nitrate release (d), inward facing with nitrite bound (e), occluded with nitrite bound (f). Once NarK completes the cycle and returns to the outward facing

conformation nitrite is released to the periplasm. The proposed mechanism is based on the rocker switch<sup>12</sup>. For NarK the rocker switch is facilitated by breaking and reforming inter-domain hydrogen bonds involving R89 and R305 as described in Fig. 2.

binding<sup>18</sup> (Fig. 3a, b). The fucose transporter FucP uses glutamates and aspartates to translocate protons together with substrate<sup>16</sup> (Fig. 3c, d). NirC, which is a proton-coupled nitrite channel has a functionally important histidine residue at the centre of the channel<sup>22</sup>. In these four examples, glutamates, aspartates and/or histidines are found on transmembrane helices with their side chains extending into the substrate translocation pathway.

NarK does not contain glutamates, aspartates or histidines on its transmembrane helices. Instead, all glutamates and aspartates are found on either cytoplasmic or periplasmic soluble domains of NarK, well away from the substrate translocation pathway (Fig. 3e, balls and sticks). Moreover, NarK only contains three histidine residues and these are also found on soluble loops. Therefore NarK has no candidate residue for proton translocation or deprotonation in its substrate translocation pathway.

Consistent with the above postulate, surface electrostatic potential calculations in NarK indicate that the substrate translocation pathway is highly positively charged. The electrostatic potentials for the N- and C-terminal halves of NarK are presented in Fig. 3f. The positively charged pathway can facilitate the transport of the negatively charged nitrate and nitrite anions, but at the same time it would represent a formidable barrier for the translocation of protons. In sharp contrast, MFS members that couple the movement of substrate to the movement of protons, have a much more balanced electrostatic distribution in their translocation pathway (Fig. 3). We note that it is possible that protons could be required for NarK activation, but our data indicate that NarK is a nitrate/nitrite exchanger in which protons are not co-transported with the substrate. Moreover, it is still not clear whether NarK could function in nitrite transport alone and if so, whether this could be bidirectional. Further functional studies would have to be performed to answer these questions.

Our data show that NarK functions as a nitrate/nitrite exchanger. We propose the following mechanism for nitrate/nitrite exchange. The mechanism is based on the structural analysis that is presented above, and the previously proposed rocker switch<sup>12</sup> (Fig. 4). Our proposed mechanism of action begins with NarK in the outward conformation in which the substrate translocation pathway is open to the periplasm (Fig. 4a). The positively charged substrate translocation pathway can attract a nitrate molecule, which can enter the pore and bind directly above the two arginines at the substrate-binding site (Fig. 4b and Supplementary Fig. 4). There the nitrate forms hydrogen bonds with R305 and N175 (Fig. 4b). The binding event could then trigger a conformational change in NarK into the transient occluded state where the pore is closed both at the periplasm and cytoplasm (Fig. 4c). The conformational change could push the nitrate from above R89 and R305 to directly below, and as the transporter adopts the inward conformation, its substrate translocation pathway opens to the cytoplasm and nitrate can then be released (Fig. 4d). As nitrate is released it is exchanged with nitrite. Nitrite enters the substrate translocation pathway and binds in plane with R89 and R305 (Figs 2d and 4e). The binding of nitrite at the substrate-binding site triggers the conformational change of NarK from its inward-facing conformation into the outward-facing conformation via the transient occluded state (Fig. 4e, f). During this process nitrite is pushed directly above R89 and R305 and once NarK is facing outward the nitrite is released into the periplasm (Fig. 4a). The cycle of exchange can then continue.

Such a rocker switch mechanism would require a hinge to allow the two halves of NarK to rock against one another as described above. We propose that the hinge is formed by the inter-domain hydrogen bonds involving the conserved arginine residues R89 and R305. As discussed above, these two residues are stabilized by inter-domain hydrogen



bonding: Y263 from the C-terminal domain and G171 from the N-terminal domain of NarK form hydrogen bonds with R305; and G144 of the N-terminal domain and S411 of the C-terminal domain of NarK form hydrogen bonds with R89 (Fig. 2c). As the substrate interacts with R89 and R305, it would disrupt the hydrogen bond networks between the two domains that are mediated by the arginines. Our structure of nitrite-bound NarK indicates that R305 undergoes a conformational change upon substrate binding consistent with a break in the inter-domain hydrogen-bonding network (Fig. 2d). We propose that it is the breaking and reforming of these inter-domain hydrogen bonds through the arginines that allow the two halves of NarK to rock against one another as depicted in Fig. 4. Consistent with this postulate, G171, G144, S411 and Y263 are residues that are conserved in NNP members (Supplementary Fig. 2).

The rocker switch mechanism proposed in other MFS members whose structures are known involves the breaking and the formation of salt bridges and hydrogen bonds between various protein residues<sup>15–19,23</sup>. As discussed above, NarK contains no acidic residues in its pore so salt bridges could only form between protein residues and the substrate, but salt bridges between various protein residues are unlikely to be involved in its mechanism of action. Instead, the pore of NarK is highly positively charged, probably to exclude protons and to attract anions like nitrate and nitrite, whereas the rocking seems to involve the breaking and formation of inter-domain hydrogen bonds at the substrate-binding pocket. It remains to be seen as more structures of MFS members from various families are determined whether other members use a similar pattern of hydrogen bond breaking and formation for their function.

## METHODS SUMMARY

NarK from *E. coli* strain K12 was overexpressed in *E. coli* BL21 (DE3) C41. Fab antibody fragments were generated as described in Methods. The NarK–Fab complex was purified in the presence of 0.2% (w/v) *n*-decyl- $\beta$ -D-maltoside and crystallized in the following condition: 0.1 M citric acid (pH 3.5), 0.1 M NaCl, 0.1 M Li<sub>2</sub>SO<sub>4</sub> and 28% PEG400. Nitrite-bound NarK–Fab crystal was obtained by soaking the NarK–Fab crystal in the buffer containing 50 mM sodium nitrite. Diffraction data sets of both crystals were collected at the Advanced Light Source (beamline 8.2.2). Data processing and structure determination were performed using the HKL2000, COOT and CCP4 programs.

**Full Methods** and any associated references are available in the online version of the paper.

Received 25 September 2012; accepted 2 April 2013.

Published online 12 May 2013.

- Wood, N. J., Alizadeh, T., Richardson, D. J., Ferguson, S. J. & Moir, J. W. Two domains of a dual-function NarK protein are required for nitrate uptake, the first step of denitrification in *Paracoccus pantotrophus*. *Mol. Microbiol.* **44**, 157–170 (2002).
- Martínez-Espinosa, R. M., Cole, J. A., Richardson, D. J. & Watmough, N. J. Enzymology and ecology of the nitrogen cycle. *Biochem. Soc. Trans.* **39**, 175–178 (2011).
- Einsle, O. & Kroneck, P. M. Structural basis of denitrification. *Biol. Chem.* **385**, 875–883 (2004).
- Saier, M. H. Jr *et al.* Phylogenetic characterization of novel transport protein families revealed by genome analyses. *Biochim. Biophys. Acta* **1422**, 1–56 (1999).

- Pao, S. S., Paulsen, I. T. & Saier, M. H. Jr. Major facilitator superfamily. *Microbiol. Mol. Biol. Rev.* **62**, 1–34 (1998).
- Jia, W. & Cole, J. A. Nitrate and nitrite transport in *Escherichia coli*. *Biochem. Soc. Trans.* **33**, 159–161 (2005).
- DeMoss, J. A. & Hsu, P. Y. NarK enhances nitrate uptake and nitrite excretion in *Escherichia coli*. *J. Bacteriol.* **173**, 3303–3310 (1991).
- Rowe, J. J., Ubbink-Kok, T., Molenaar, D., Konings, W. N. & Driessen, A. J. NarK is a nitrite-extrusion system involved in anaerobic nitrate respiration by *Escherichia coli*. *Mol. Microbiol.* **12**, 579–586 (1994).
- Moir, J. W. & Wood, N. J. Nitrate and nitrite transport in bacteria. *Cell. Mol. Life Sci.* **58**, 215–224 (2001).
- Jia, W., Tovell, N., Clegg, S., Trimmer, M. & Cole, J. A single channel for nitrate uptake, nitrite export and nitrite uptake by *Escherichia coli* NarU and a role for NirC in nitrite export and uptake. *Biochem. J.* **417**, 297–304 (2009).
- Wang, Y. Y., Hsu, P. K. & Tsay, Y. F. Uptake, allocation and signaling of nitrate. *Trends Plant Sci.* **17**, 458–467 (2012).
- Law, C. J., Maloney, P. C. & Wang, D. N. Ins and outs of major facilitator superfamily antiporters. *Annu. Rev. Microbiol.* **62**, 289–305 (2008).
- Sun, L. *et al.* Crystal structure of a bacterial homologue of glucose transporters GLUT1–4. *Nature* **490**, 361–366 (2012).
- Solcan, N. *et al.* Alternating access mechanism in the POT family of oligopeptide transporters. *EMBO J.* **31**, 3411–3421 (2012).
- Newstead, S. *et al.* Crystal structure of a prokaryotic homologue of the mammalian oligopeptide-proton symporters, PepT1 and PepT2. *EMBO J.* **30**, 417–426 (2011).
- Dang, S. *et al.* Structure of a fucose transporter in an outward-open conformation. *Nature* **467**, 734–738 (2010).
- Yin, Y., He, X., Szweczyk, P., Nguyen, T. & Chang, G. Structure of the multidrug transporter EmrD from *Escherichia coli*. *Science* **312**, 741–744 (2006).
- Huang, Y., Lemieux, M. J., Song, J., Auer, M. & Wang, D. N. Structure and mechanism of the glycerol-3-phosphate transporter from *Escherichia coli*. *Science* **301**, 616–620 (2003).
- Abramson, J. *et al.* Structure and mechanism of the lactose permease of *Escherichia coli*. *Science* **301**, 610–615 (2003).
- Trueman, L. J., Richardson, A. & Forde, B. G. Molecular cloning of higher plant homologues of the high-affinity nitrate transporters of *Chlamydomonas reinhardtii* and *Aspergillus nidulans*. *Gene* **175**, 223–231 (1996).
- Mirza, O., Guan, L., Verner, G., Iwata, S. & Kaback, H. R. Structural evidence for induced fit and a mechanism for sugar/H<sup>+</sup> symport in LacY. *EMBO J.* **25**, 1177–1183 (2006).
- Lü, W. *et al.* Structural and functional characterization of the nitrite channel NirC from *Salmonella typhimurium*. *Proc. Natl Acad. Sci. USA* **109**, 18395–18400 (2012).
- Qin, L. *et al.* Sialin (SLC17A5) functions as a nitrate transporter in the plasma membrane. *Proc. Natl Acad. Sci. USA* **109**, 13434–13439 (2012).
- Harlow, E. & Lane, D. *Antibodies: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, 1988).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank E. McCleskey for critically reading this manuscript and for discussions. We thank D. Cawley for development and production of monoclonal antibodies, and staff at the Advanced Light Source, Lawrence Berkeley National Laboratory for assistance with X-ray data collection. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the US Department of Energy under Contract no. DE-AC02-05CH11231. Research in the Gonen laboratory is funded by the Howard Hughes Medical Institute.

**Author Contributions** H.Z. and T.G. designed the project. H.Z. performed all biochemical experiments including cloning, expression, purification, antibody production and binding assays, crystallization and X-ray data collection for both apo- and nitrite-bound NarK. H.Z. and G.W. built and refined the structures. All authors participated in data analysis and figure preparation. H.Z. and T.G. wrote the manuscript.

**Author Information** Structures of substrate-free and nitrite-bound NarK have been deposited in PDB under accession numbers 4JR9 and 4JRE, respectively. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.G. ([gonent@janelia.hhmi.org](mailto:gonent@janelia.hhmi.org)).

## METHODS

**Protein expression and purification.** The gene encoding full-length NarK from *E. coli* strain K12 was subcloned into pET15b (EMD Millipore) with a modified N-terminal 8× His-tag and a thrombin cleavage site. NarK was overexpressed in *E. coli* BL21 (DE3) C41 at 37 °C for 4–5 h with 0.3 mM IPTG as inducer. After induction, the cells were collected by centrifugation and resuspended in lysis buffer containing 20 mM Tris-HCl (pH 8), 150 mM NaCl and 1 mM PMSF. The resuspension was then passed through a microfluidizer (Microfluidics Corporation) twice at ~15,000 p.s.i., followed by centrifugation at 15,000g for 30 min. The supernatant was collected and centrifuged at 130,000g for 1 h. The pellets containing membrane were resuspended in the same lysis buffer and frozen at –80 °C until use.

To purify NarK, an aliquot of frozen membrane was thawed and solubilized with 1% *n*-decyl- $\beta$ -D-maltopyranoside (DM) at 4 °C for 2 h. After addition of 20 mM imidazole followed by centrifugation, the supernatant was applied to Ni<sup>2+</sup> nitrilotriacetate affinity resin (Ni-NTA). The resins were washed with 50 mM imidazole in the buffer containing 20 mM Tris-HCl (pH 8), 150 mM NaCl, 0.2% DM. Full-length NarK was then eluted with 250 mM imidazole in the same buffer. The His-tag was removed by thrombin digestion at an enzyme:protein molar ratio of 1:1,000 at 4 °C overnight. The enzyme-treated protein was further purified by gel filtration (Superdex-200) in 20 mM Tris-HCl (pH 8), 150 mM NaCl and 0.2% DM. NarK in the peak fractions was collected.

**Fab production.** Standard protocol<sup>24</sup> was used to generate the mouse IgG monoclonal antibodies against NarK. Western blot and native ELISA was performed to assess the binding affinity and specificity of the antibodies generated from hybridoma cell lines. Several monoclonal antibodies with strongest binding affinity were then purified from the hybridoma supernatants by protein A affinity chromatography. Fab was produced by papain digestion and purified by protein A affinity chromatography.

**Assembly of NarK–Fab complex.** Purified NarK and Fab were mixed at a molar ratio of 1:2, and incubated at 4 °C for 30 min. The complex was then concentrated and purified by gel filtration chromatography with 20 mM Tris-HCl (pH 8), 150 mM NaCl and 0.2% DM. A clear and complete peak shift to higher molecular weight was observed, indicating homogeneous NarK–Fab complex formation. The purified protein complex was collected and concentrated to 5.8 mg ml<sup>–1</sup>.

**Crystallization.** Initial hanging-drop crystallization assay with purified NarK produced crystals grown in a large range of crystallization condition with PEG molecules. However, these crystals gave anisotropic diffraction to around 4 and 6 Å. High-quality crystals were obtained only when NarK was co-crystallized as a complex with Fab. The best crystal, which diffracted to ~2.5 Å, was obtained with Fab prepared from hybridoma line 4G5 (IgG2a, kappa) in the following precipitant condition: 0.1 M citric acid (pH 3.5), 0.1 M NaCl, 0.1 M Li<sub>2</sub>SO<sub>4</sub> and 28% PEG400. Before data collection, the crystals were soaked in a cryoprotectant buffer containing 30% PEG400 in the same precipitant solution for 5 min, and rapidly

frozen in liquid nitrogen. For soaking, the NarK crystals were transferred into a precipitant solution containing 50 mM sodium nitrite or sodium nitrate overnight, and then frozen as above. Nitrite-bound-NarK crystals diffracted to ~2.8 Å resolution.

**Data collection and structure determination.** The data sets were collected at the Advanced Light Source (beamline 8.2.2), and processed with HKL2000<sup>25</sup> to 2.6 Å (substrate-free NarK) and 2.8 Å (nitrite-bound NarK) resolution. Further structure determination and refinement were accomplished using the CCP4 package<sup>26</sup>.

The structure was determined by molecular replacement using the program Phaser<sup>27</sup> with a polyaniline Fab fragment derived from the Protein Data Bank (PDB ID 1F8T<sup>28</sup>) as a search model. Phases from molecular replacement were significantly improved after cycles of density modification using the program Parrot<sup>29</sup>, and the electron density for the 12 transmembrane helices became apparent. Manual model building was carried out for the Fab using the program Coot<sup>30</sup>, followed by structure refinement using the program Refmac<sup>31</sup>. To facilitate model building of NarK, 20-residue polyaniline helices were placed in the asymmetric unit by molecular replacement. Subsequent cycles of density modifications, model building and refinement were carried out until structure completion. The final model contains one molecule of NarK (residues 12–458) and one molecule of the heavy and light chains of the Fab in the asymmetric unit. Data collection and refinement statistics are presented in Supplementary Table 1.

All figures in this paper were prepared with Chimera version 1.6.2<sup>32</sup> or Pymol version 1.5<sup>33</sup> and assembled in Photoshop CS6 (Adobe). Supplementary Fig. 2 was prepared using the program Clustal X<sup>34</sup>.

25. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
26. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67**, 235–242 (2011).
27. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
28. Fokin, A. V. *et al.* Spatial structure of a Fab-fragment of a monoclonal antibody to human interleukin-2 in two crystalline forms at a resolution of 2.2 and 2.9 angstroms [in Russian with English abstract]. *Bioorg. Khim.* **26**, 571–578 (2000).
29. Zhang, K. Y., Cowtan, K. & Main, P. Combining constraints for electron-density modification. *Methods Enzymol.* **277**, 53–64 (1997).
30. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
31. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
32. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
33. The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC <http://www.pymol.org/>.
34. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).

## ERRATUM

doi:10.1038/nature12196

### Erratum: Basic amino-acid side chains regulate transmembrane integrin signalling

Chungho Kim, Thomas Schmidt, Eun-Gyung Cho, Feng Ye, Tobias S. Ulmer & Mark H. Ginsberg

*Nature* **481**, 209–213 (2012); doi:10.1038/nature10697

The legend to Fig. 1d of this Letter incorrectly referred to “a POPS lipid’s amino  $\text{NH}_3^+$  group (blue)”. It should read “a POPC lipid’s choline  $\text{N}(\text{CH}_3)_3^+$  group (blue)”, where POPC is 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine. The original legend to Fig. 1d has been corrected.

## RETRACTION

doi:10.1038/nature12164

### Retraction: Branched tricarboxylic acid metabolism in *Plasmodium falciparum*

Kellen L. Olszewski, Michael W. Mather, Joanne M. Morrissey, Benjamin A. Garcia, Akhil B. Vaidya, Joshua D. Rabinowitz & Manuel Llinás

*Nature* **466**, 774–778 (2010); doi:10.1038/nature09301 and corrigendum *Nature* **469**, 432 (2011); doi:10.1038/nature09712

We retract this Letter, which reported both reductive and oxidative tricarboxylic acid (TCA) metabolism in *Plasmodium falciparum* parasites (‘branched TCA metabolism’). The data for metabolic labelling of TCA intermediates remain reliable, but we have come to realize that the interpretation presented is incorrect. Although there is both reductive and oxidative TCA cycle flux in *P. falciparum*-infected red blood cell (RBC) cultures (as we reported), new data from the Llinás and Vaidya groups (manuscript in preparation) suggests that the reductive flux occurs primarily in the RBCs and not in the parasite itself. Specifically, we have used new enrichment strategies for the parasitized RBCs that enhance our ability to measure *P. falciparum*-infected RBC metabolic activity without excessive interference from surrounding uninfected RBCs. On feeding  $^{13}\text{C}_5$  glutamine, we measured both  $^{13}\text{C}_2$ -succinate and  $^{13}\text{C}_4$ -succinate in the infected RBCs, demonstrating that TCA metabolism in blood-stage *P. falciparum* is not branched but primarily oxidative (cyclic).



# CAREERS

**CHEMISTRY** US National Research Council starts work on lab-safety guidelines **p.655**

**IMMIGRATION** Phone scam targets foreign postdocs in United States **p.655**

**NATUREJOBS** For the latest career listings and advice [www.naturejobs.com](http://www.naturejobs.com)

ALEKSANDER1/SHUTTERSTOCK



## TECHNOLOGIES

# The right tool

*Research-tool developers must be creative, innovative and willing to collaborate with people from a variety of fields.*

BY KELLY RAE CHI

Gregory Buckner once dreamed of being a surgeon. But his father, a civil engineer, convinced him to try engineering. Now, as a mechanical engineer at North Carolina State University in Raleigh, Buckner combines the two interests. He creates, builds and develops biomedical tools such as

robotic catheters for use in minimally invasive cardiothoracic surgery. Buckner does not work directly with chest pain or lacerations. Instead, he deals with batteries and circuit boards and plastic tubing.

Tools from lab instruments to software can refine research and open new fields of enquiry. And the people who make them are a varied bunch — they might be full-time university

researchers, or they might be hired by larger teams in academia or industry. The boundaries are fluid. Buckner, for example, has worked as a consultant, licensed his inventions to industry and even launched his own company.

Industry jobs involve specific goals with fast turnarounds. They are often at companies that sell scientific instrumentation, medical devices or software, such as Life Technologies, based in Carlsbad, California; Oxford Instruments in Abingdon, UK; and Medtronic in Minneapolis, Minnesota. Options are many for those keen on focusing their skills and scientific know-how on tools rather than lab research.

“There’s certainly a diversity of ways to impact science through tool building,” says Eric Betzig, a group leader at the Howard Hughes Medical Institute’s Janelia Farm Research Campus in Ashburn, Virginia, who at Janelia and elsewhere has led the development of microscopes that can image features smaller than half the wavelength of the light used to illuminate them. Often overlooked in the past, designing and building biomedical tools has been receiving growing respect as a career track. There are now grants and publications dedicated to the craft, and toolmakers are increasingly integrated into research teams to help not only with the development of tools, but also with their application. Although they may not be responsible for the actual findings, toolmakers can help to overcome research obstacles in the lab or the clinic.

## ALL WALKS OF LIFE

Successful tool builders typically have expertise in fields such as engineering, computer science, bioinformatics, maths or physics — and occasionally biology. They also tend to have a willingness to learn, a knack for picking up mechanical and research skills and enthusiasm for collaboration. Members of Janelia Farm’s instrument design and fabrication group, for example, have worked in the defence and automotive industries. They learn neuroscience on the job at Janelia, says Chris Werner, the group’s shared-resource director. Tool builders shadow scientists in the laboratory until they understand the problem at hand, and they may even be named as co-authors on publications. “Besides technical skills, what’s really important and makes the difference is their soft skills — communication, giving credit, taking credit — the ability to play nice,” says Werner.

Toolmakers need to be good listeners, ►

► alert to the challenges of bench research — whether those are imaging a live cell or analysing a huge number of data. Developing tools to solve such problems involves making multiple test runs and repeatedly conferring with users or customers to make sure that the instrument or algorithm is doing what it needs to. This can be a big challenge — often researchers are unsure of exactly what they need. And success can mean being anchored to a project for a long time; tool builders frequently get roped into collaborations and companies centred on their technologies.

Armon Sharei, a PhD student in chemical engineering at the Massachusetts Institute of Technology in Cambridge, saw at first hand the healthy dose of trial and error that can be involved in tool development. Sharei embarked on the path to an invention by chance, when he was trying to shoot cells with a jet of liquid to deliver proteins, DNA, nanoparticles and other large molecules into the cells for research and clinical applications. The delivery worked, but not for the reasons that he expected. Instead of directly injecting the membrane with fluid, the jet seemed to be squashing the cells — squeezing them so much that it caused their membranes to open temporarily and allow the fluid to diffuse in. But delivery efficiency was poor. “The prospects didn’t look good but what kept us going is that, in principle, this method was so simple it could make a big difference if we could improve it,” says Sharei.

His research advisers took a risk and the group overhauled the design. The team eliminated the jet and created different versions of a microfluidic chip that worked by essentially squeezing the cells. The team patented the technology and in March launched SQZ Biotech in Boston, Massachusetts, to commercialize the device. Applications might include delivering reprogramming proteins to convert fully developed cells into induced pluripotent stem cells. “I think that lucky first observation or experiment comes to everyone once in a while. It’s just a matter of recognizing what it is and developing its potential,” says Sharei.

Serendipity aside, tool ideas often come from a deep understanding of customers’ needs, which provides insight into ways to make improvements. Ideas for most of Buckner’s developments — which include a finely controlled robotic catheter and a chest

retractor that measures force during open-heart surgery — come from watching surgeons, cardiologists and radiation oncologists. “They are true experts in the field and know what the technical hurdles of their current practices are,” says Buckner, who has patents on multiple inventions.

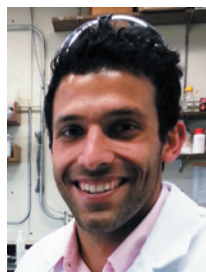
Tool development itself often spawns further ideas; one project may proliferate into many. Through a clinical collaboration in 2007–08, Buckner’s group was working to automate the tying of knots in sutures during surgery on the mitral valve of the heart. The researchers had to test their prototypes on live pigs, which proved both costly (at US\$2,500 per pig) and time consuming. So they developed a ‘dynamic heart system’ that pumped fluid through a dead pig heart so that it functioned realistically. The device allowed the team to test its technologies for less than \$25 a go.

## SILVER LINING

Toolmakers must be ready for some disappointment — and if the end product fails to live up to the original idea, they should be nimble enough to tweak it, find other applications or move on. Betzig experienced difficulties with near-field microscopy, a high-resolution imaging method. The technique turned out to have some physical limitations that made it less than ideal for its intended use in imaging biological samples. In the end, he moved on to different projects, although others have continued to pursue near-field microscopy. “The thing about tools is that they’re kind of like your children,” says Betzig. “When they’re born, you say, ‘Oh, he could be president! Or an astronaut! Cure cancer!’” But toolmakers often come to the realization that their work might not have quite the impact that they anticipated. Betzig says that his tools have failed in various ways — and that by focusing on the flaws, he can find opportunities to develop better versions.

Once a tool builder based in academia believes that he or she has something ready for the marketplace, another set of challenges awaits. In the United States, for example, the developer must often file an invention disclosure with his or her university’s technology-transfer office, allowing the office to evaluate the potential for commercializing the tool. If it looks promising, the office then usually files a provisional patent application. If certain conditions specified by the US Patent and Trademark Office are met within 12 months, the application can be made non-provisional.

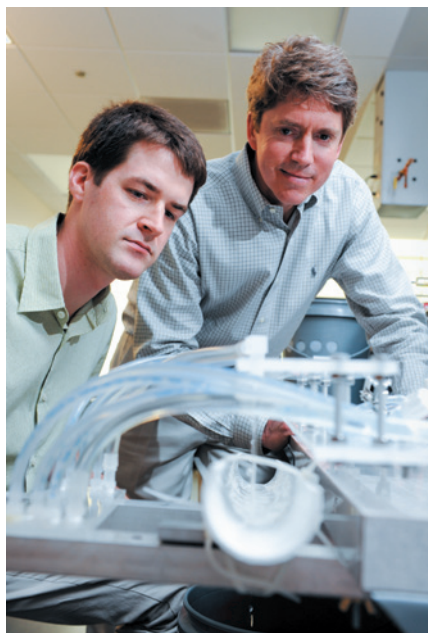
After a patent is issued, there are numerous possibilities. The inventor might seek funds from industry or the government to refine the prototype. The technology-transfer office might work with the inventor to find companies that are willing to license it. Any tools geared for clinical use will need to be submitted for regulatory approval. And the inventor could launch a company with the help of partners or



*“That lucky first observation or experiment comes to everyone. It’s just a matter of recognizing it and developing its potential.”*

Armon Sharei





Mechanical engineer Gregory Buckner (right) works with a graduate student.

advisers who have business and legal expertise. After tools are sold, new applications might come to light, and the company can develop these or create spin-offs to do so.

### BUILDING SKILLS

Just like researchers, toolmakers can learn through an apprenticeship of sorts. “As a more-junior graduate student, one can learn these skills very deeply by working with more-senior graduate students, postdocs and principal investigators,” says Samuel Hess, a physicist at the University of Maine in Orono who has invented high-resolution imaging methods. Degrees in fields such as biomedical engineering or computational biology can help people to build skills at the intersection of toolmaking and medicine or biology. Even after obtaining their PhDs, toolmakers can learn or refine their skills by working with more-experienced researchers during a postdoctoral fellowship, on sabbatical or in industry. Whenever it happens, learning to build is a hands-on process — with a lot of mistakes made along the way, says Betzig.

Many established toolmakers are essentially self-taught and self-motivated. “I hire lots of programmers and engineers. The best programmers are always former biologists who pick up the programming bug,” says Chris Beecher, an analytical chemist and co-founder of IROA Technologies in Ann Arbor, Michigan, who has invented platforms that quantify the small molecules produced by metabolism in cells or tissues.

At Waters Corporation, an analytical science instrumentation company based in Milford, Massachusetts, some of the product developers are not engineers, but scientists who have previous experience using analytical

chemistry instruments as part of their research, so they know what customers need, says Steven Cohen, life-sciences director in research and development at the company. Such developers work with a manufacturing workshop to design instruments and build prototypes.

In academia and the non-profit sector, tool development is often funded through a larger research project. But researchers can also apply for grants specific to tool building. Genome Canada in Ottawa, for example, supports the development of genomic technologies through funding competitions. Naveed Aziz, the organization’s director of technology programmes, says that he receives funding applications led by bioinformaticians as often as by researchers. And the US National Science Foundation (NSF) runs an Instrument Development for Biological Research grant to support development of tools that enable new research capabilities, markedly improve current technology or transform a prototype into something more broadly usable.

Several US federal agencies — the NSF, the National Institutes of Health and the Department of Defense — also offer Small Business Innovation Research and Small Business Technology Transfer grants, which aim to spur technological innovation and lower the barriers to commercialization. The Wellcome Trust in London, one of the world’s biggest biomedical research charities, uses its Translation Fund and Health Innovation Challenge Fund to speed the commercialization of biomedical technologies.

Grants like these can help toolmakers make a lasting difference to a given field. It is always nice to go “beyond the narrow scope” of the original project, notes Charles Schmitt, director of informatics at the Renaissance Computing Institute, based in Chapel Hill, North Carolina. But although it is gratifying when a tool is adopted widely, it sometimes means that toolmakers must train others to use their inventions, and must surrender control to groups with their own skill sets and agendas. Toolmakers can take pride in their impact, but lose command of their brainchild. Moving on, says Schmitt, “is always a big challenge.” ■

**Kelly Rae Chi** is a freelance writer based in Cary, North Carolina.



**“One can learn these skills by working with more-senior graduate students, postdocs and principal investigators.”**

Samuel Hess

### CHEMISTRY

## Improving lab safety

A US National Research Council committee met on 15–16 May in Washington DC to start developing safety recommendations for chemical researchers in academic and national laboratories. The project was spurred in part by three high-profile accidents at US academic laboratories between 2008 and 2011. Behavioural scientists, chemists and safety experts will identify problems at research institutions and learn from effective safety systems in industry to develop the recommendations, which will be released next spring. Chemists and behavioural scientists “can learn from each other,” says committee chairman Holden Thorp, chancellor of the University of North Carolina at Chapel Hill.

### BRAIN DRAIN

## Workers flee corruption

A growing number of highly skilled workers including researchers are leaving corrupt nations where government officials demand bribes and control access to the labour market, finds a study published on 17 May (A. Ariu and M. P. Squicciarini *EMBO Rep.* <http://doi.org/mkh>; 2013). Nations with relatively low corruption benefit from an influx of scientists who write influential papers and patents and create businesses, says the study, which examined movement patterns in 123 nations against an international corruption index. “It is not a positive thing for a researcher to be in a country that is highly corrupt,” says study co-author Mara Squicciarini, an economist at the Catholic University of Leuven in Belgium.

### IMMIGRATION

## Postdocs hit by scam

Confidence tricksters are targeting international postdocs and students in the United States, according to complaints passed on to the National Postdoctoral Association (NPA) in Washington DC. Telephone callers claiming to be from US Citizenship and Immigration Services (USCIS) tell trainees that they have improperly completed a form, and that they face a fine or deportation. This is not agency procedure, says the NPA. Callers know the target’s name, date of birth, address, phone number and case number. People who suspect that they have been targeted should report the scam at [go.nature.com/poexjx](http://go.nature.com/poexjx). The USCIS did not respond to requests for comment.



# A TIME FOR PEACE

*Future plans.*

BY S. R. ALGERNON

“Why do you deserve a place here, Oscar?” asked the director, as he eyed Oscar from across a sturdy oak desk. A painting of Speer’s *Volkerbundhalle* and the Berlin skyline dominated the wall behind him.

Oscar looked down at his fidgety knees to avoid the director’s glare. He had not prepared himself for silence. He had imagined the Institute as a place where the air danced with the roar of engines and the crackle of a welder’s torch, or at least the clack of chalk against slate.

Nobody deserved a place at the Institute, thought Oscar. The Institute plucked them from the world like the tip of a tornado and hurled them into the maelstrom of history. Fermi put his studies on hold for three years and emerged in 1924 with a complete understanding of nuclear fission. Alan Turing took a summer off from Sherborne and crossed the pond to New Jersey. By 1935, the first business computers rolled off the production line.

Ten seconds passed. The director’s office was quiet, except for the steady sweeping of a broom and the snip of hedge trimmers outside the window.

Oscar swallowed and took a breath.

“It all comes down to quantum tunnelling. We’ve always assumed that time travel involved moving through space-time in accordance with Dr Einstein’s theories, but if Dr Born is right...”

“It is not a question of *if*,” said the director. He reached into his desk and pulled out a photograph. “This was our founding conference, after the Great War. Look in the third row.”

An inscription in the upper left corner read POTSDAM, 1919. Just below it, between Heisenberg and Rutherford, Oscar saw a bearded version of himself.

“I was there?”

“You and your boxcar full of journals and blueprints. You convinced Heisenberg, and he convinced the rest. Together, you drew up the plans for the twentieth century.”

“So... I am, or will be, the first time traveller?”

“The Oscar that appeared in Potsdam 50 years ago is a matter of historical record. I want to know who *you* are. Are you — the you that sits in my office — necessary to move history onto its present path?”

“The grandfather paradox,” said Oscar,

with a chuckle. “It’s not a problem. Moving through time implies velocity and causation. Quantum tunnelling through time is different. The time traveller simply appears *ex nihilo* from the quantum foam. The traveller’s emergence on the scene is, to be sure, an improbable event, but it has no destiny and it owes nothing to the future.”



“So, you do not have a destiny?”

“Maybe not,” said Oscar, “but I hope you’ll agree that I still have potential. I will find a way to unwind time to an earlier state. The history books tell us that Brown and Reitsch walked on the Moon in 1947. What if it were Lindbergh in 1927 or da Vinci in 1497?”

The director stroked his chin.

“And what about all the people in this timeline? Will they agree to wink out of existence to create your perfect world?”

“Once the timeline is unwound, everyone here will be unrealized potential. They won’t get a choice. After all, nobody gets to vote on what time it is.”

“But you do. Or, you will. Doesn’t that make you responsible for the people you leave behind?”

“What can I do for them? They won’t even exist. Am I obliged to consider every potential human life, not just every embryo but every possible union of sperm and egg?”

“Perhaps not,” said the director. “I think we have a place for you after all. Our programme will be

intensive. Have you explained to your friends and relatives that you will be quite busy for some time and should not be disturbed?”

“Of course.”

“Good,” said the director. He stood up and offered his hand. “Welcome. I will show you to your room.”

The room, at the other end of the quad, was sparse and tidy, like a monk’s cell. A grey jumpsuit lay neatly on the bed. At first, the sight startled him. Did Turing have to sweep the paths when he was here? Did Fermi have to trim the hedges?

Oscar stepped tentatively up to the window. Outside, a man sat with his back to Oscar, facing a watercolour of a Bavarian castle against the backdrop of an idyllic Wagnerian landscape. Oscar tried to lean out the window to get the artist’s attention, but it opened only a few centimetres, not enough for him to squeeze his head through.

“Hallo, *nachbar*,” said Oscar, eager to try out a little German.

The artist slowly turned, without lowering his brush.

“Ja,” said the painter, without inflection, looking vaguely towards Oscar’s window. “Hallo, *nachbar*.” His head drifted back towards the canvas, and he added a few brushstrokes.

Oscar took a step back, unimpressed. He had expected to meet the next Einstein or Heisenberg, not some undersized Charlie Chaplin. He turned back to the director.

“There must be some mistake,” said Oscar. “I’m no painter or gardener. What about my future? What about history?”

The old man’s smile drew the skin taut over his face.

“You may owe this world no debt, but we have our own obligations. I’m sure you’ll understand if we prefer to leave the timeline as it is.”

Oscar stepped towards the door, but the director slammed it shut. On the other side, a deadbolt slid firmly into place.

Oscar staggered backwards to the bed, hearing the sound of his own quickened heartbeat and the even cadence of pruning shears. After each slice, branches struck the pavement and waited for the broom to sweep them away. ■

**S. R. Algernon** studied fiction writing and biology, among other things, at the University of North Carolina at Chapel Hill. He currently lives in Singapore.

➔ **NATURE.COM**  
Follow Futures:  
@NatureFutures  
go.nature.com/mtoodm

# The role of epistasis in protein evolution

ARISING FROM M. S. Breen, C. Kemena, P. K. Vlasov, C. Notredame & F. A. Kondrashov *Nature* **490**, 535–538 (2012)

An important question in molecular evolution is whether an amino acid that occurs at a given site makes an independent contribution to fitness, or whether its contribution depends on the state of other sites in the organism's genome, a phenomenon known as epistasis<sup>1–5</sup>. Breen and colleagues recently argued<sup>6</sup> that epistasis must be “pervasive throughout protein evolution” because the observed ratio between the per-site rates of non-synonymous and synonymous substitutions (dN/dS)<sup>7</sup> is much lower than would be expected in the absence of epistasis. However, when calculating the expected dN/dS ratio in the absence of epistasis, Breen *et al.*<sup>6</sup> assumed that all amino acids observed at a given position in a protein alignment have equal fitness. Here, we relax this unrealistic assumption and show that any dN/dS value can in principle be achieved at a site, without epistasis; furthermore, for all nuclear and chloroplast genes in the Breen *et al.* data set, we show that the observed dN/dS values and the observed patterns of amino-acid diversity at each site are jointly consistent with a non-epistatic model of protein evolution.

For a variety of proteins under purifying selection, Breen *et al.*<sup>6</sup> constructed alignments and recorded the amino acids observed at each position; these observed amino acids were deemed “acceptable” with respect to natural selection. They then assumed that substitutions occur at neutral rates among the acceptable amino acids in order to calculate, for each protein, an expected value for dN/dS in the absence of epistasis. Because their empirical observations of dN/dS were much lower than these expected values, Breen *et al.*<sup>6</sup> concluded that epistasis must be extremely prevalent.

The flaw in this reasoning is that Breen *et al.*<sup>6</sup> considered only a single class of fitness assignments, so that all amino acids observed at a site were assumed equally fit. A more realistic assumption is that some amino acids observed at a site are more fit than others<sup>8,9</sup>.

To illustrate the principle that low dN/dS can arise without epistasis, we considered a non-epistatic model in which, among the acceptable amino acids at a given site, one of these is preferable to the rest. We performed the following experiment: in a hypothetical protein of length 300 amino acids, for each position we randomly designated eight amino acids as acceptable (the average number of acceptable amino acids reported by Breen *et al.*<sup>6</sup>), but gave one of these a selective advantage over the rest. We then calculated the equilibrium dN/dS (ref. 10) for this protein as a function of the selective advantage of the preferred amino acid,  $2Ns$  (Fig. 1). Whereas dN/dS is high for the case  $2Ns = 0$ , corresponding to the Breen *et al.*<sup>6</sup> assumption, dN/dS is much lower for larger  $2Ns$ . Thus, a large range of dN/dS values are consistent with non-epistatic models of protein evolution.

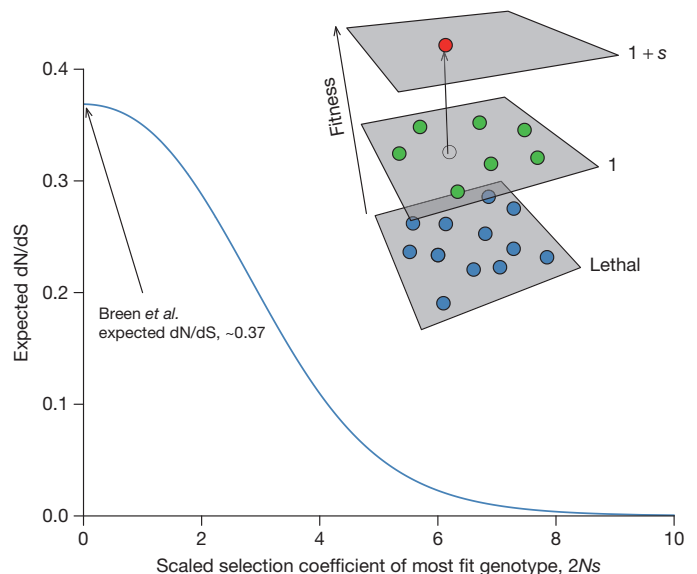
Although non-epistatic models can in principle produce low dN/dS values (Fig. 1), can such a model account for the Breen *et al.*<sup>6</sup> data? To answer this question, we considered a more general non-epistatic model that assigns to each amino acid at a site a different fitness. For each gene in the Breen *et al.*<sup>6</sup> data set, we assigned fitnesses at each site in such a way that the resulting equilibrium distribution of amino acids under our model precisely matches the amino-acid frequencies observed for that site<sup>11</sup>. Furthermore, as a result of these fitness assignments, the asymptotic mean pairwise sequence divergence under our model necessarily matches the mean pairwise divergence observed in the data (Table 1 and Methods).

Using this model, for each gene in the Breen *et al.*<sup>6</sup> data set we repeatedly simulated the evolution of a pair of sequences from their common ancestor and computed dN/dS. For the 13 mitochondrial genes, the average simulated dN/dS values, although substantially

lower than the Breen *et al.*<sup>6</sup> expectations, are still greater than the empirically observed values (Table 1). However, for the three nuclear and chloroplast genes in the Breen *et al.*<sup>6</sup> data set, the average dN/dS values under our non-epistatic model are comparable to or even lower than the empirical dN/dS values Breen and colleagues reported. Thus, the dN/dS values observed in these genes need not be attributed to epistasis, but rather can be explained by the more parsimonious assumption that the various amino acids observed at a site have different fitnesses.

It is important to note that the effects of natural selection and phylogeny are confounded in the amino-acid frequencies observed at each site, and therefore in our fitness estimates. Although methods exist to disentangle these effects when the phylogeny is small and known<sup>12,13</sup>, there is no well-accepted phylogeny for the vast range of taxa studied by Breen *et al.*<sup>6</sup>. Nonetheless, whatever the true phylogeny may be, under the standard assumption that molecular evolution can be modelled as an equilibrium Markov chain (see, for example, ref. 14, as used by Breen *et al.*<sup>6</sup>) our fitness estimates are maximum likelihood. Relaxing this assumption, or allowing more complex models (for example, allowing fitnesses or population sizes to vary across time or clade), would make it only more difficult to reject the non-epistatic null hypothesis.

In summary, Breen *et al.*<sup>6</sup> provide no direct evidence of epistasis, nor do they reject the full space of non-epistatic models. They have analysed only three non-mitochondrial genes, whose evolutionary patterns, we have shown, can be explained without epistasis. Although Breen *et al.*<sup>6</sup> contend that epistasis is the primary factor in all of molecular evolution, further work is needed to substantiate this claim.



**Figure 1 | Non-epistatic models of protein evolution can produce low dN/dS values.** Expected dN/dS as a function of  $2Ns$  for a hypothetical protein of length 300, in which eight acceptable amino acids are chosen at random for each position and one of these amino acids at random is assigned a selective advantage of size  $2Ns$ . The remaining 12 amino acids are lethal. The Breen *et al.*<sup>6</sup> expectation for dN/dS in the absence of epistasis corresponds to  $2Ns = 0$ .

Table 1 | Observed and expected dN/dS values

Gene	Breen <i>et al.</i> <sup>6</sup> expected dN/dS	Our average simulated dN/dS	Breen <i>et al.</i> <sup>6</sup> empirical dN/dS	Our equilibrium mean pairwise divergence	Breen <i>et al.</i> <sup>6</sup> empirical pairwise divergence
Mitochondrial					
ATP6	0.44	0.215	0.056	0.332	0.332
ATP8	0.56	0.624	0.224	0.615	0.615
COX1	0.28	0.078	0.015	0.188	0.188
COX2	0.43	0.140	0.025	0.348	0.348
COX3	0.32	0.144	0.036	0.290	0.290
CYTB	0.51	0.117	0.039	0.242	0.242
ND1	0.39	0.208	0.040	0.383	0.383
ND2	0.51	0.262	0.067	0.398	0.398
ND3	0.49	0.242	0.069	0.379	0.379
ND4	0.42	0.239	0.045	0.433	0.433
ND4L	0.49	0.369	0.076	0.502	0.502
ND5	0.32	0.211	0.057	0.407	0.407
ND6	0.42	0.397	0.073	0.554	0.554
Nuclear					
EEF1A1	0.11	0.031	0.020	0.080	0.080
H3.2	0.14	0.014	0.037	0.019	0.019
Chloroplast					
rbcL	0.40	0.024	0.072	0.056	0.056

Comparison of expected dN/dS values and mean pairwise divergence with the empirical values for each gene in the Breen *et al.*<sup>6</sup> data set. The Breen *et al.* expected dN/dS is based on the assumption that all amino acids observed at a given site are neutral relative to each other. Our expected dN/dS is based on the assumption that the various amino acids observed at a site have different fitnesses.

METHODS

We assume that each codon evolves according to an independent Markov chain, the rate matrix of which is determined by the scaled selection coefficient assigned to each amino acid<sup>15</sup>. The equilibrium frequency of each amino acid is then proportional to  $u_i e^{2Ns(i)}$  (ref. 15), in which  $u_i$  is the number of codons that code for amino acid  $i$ , and  $2Ns(i)$  is its scaled selection coefficient. After assigning site-specific fitnesses to amino acids, 1,000 simulations were conducted for each protein, as follows. For each site represented in at least half the sequences from the Breen *et al.*<sup>6</sup> alignment, an ancestral codon was drawn from the equilibrium distribution of our Markov chain. Two copies of this ancestral sequence were then evolved independently until dS = 0.25, which is within the range of dS = 0.05 to 0.5 used by Breen *et al.*<sup>6</sup>. We then estimated dN/dS for each pair using PAML<sup>14</sup>, again following the procedure of Breen *et al.* Mean pairwise divergence (Table 1) was calculated using the formula  $\frac{1}{L} \left( \sum_{j=1}^L \left( 1 - \sum_{i=1}^{20} f_{ij}^2 \right) \right)$ , in which  $f_{ij}$  denotes the frequency of amino acid  $i$  at site  $j$ , and  $L$  the number of majority non-gapped sites in the protein. All computer code is available on request.

David M. McCandlish<sup>1</sup>, Etienne Rajon<sup>1</sup>, Premal Shah<sup>1</sup>, Yang Ding<sup>1</sup> & Joshua B. Plotkin<sup>1</sup>  
<sup>1</sup>Department of Biology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA.  
e-mail: jplotkin@sas.upenn.edu

Received 20 December 2012; accepted 18 April 2013.

1. Kondrashov, A. S., Sunyaev, S. & Kondrashov, F. A. Dobzhansky-Muller incompatibilities in protein evolution. *Proc. Natl Acad. Sci. USA* **99**, 14878–14883 (2002).

2. DePristo, M. A., Weinreich, D. M. & Hartl, D. L. Missense meanderings in sequence space: a biophysical view of protein evolution. *Nature Rev. Genet.* **6**, 678–687 (2005).

3. Kryazhimskiy, S., Dushoff, J., Brazykin, G. A. & Plotkin, J. B. Prevalence of epistasis in the evolution of influenza A surface proteins. *PLoS Genet.* **7**, e1001301 (2011).

4. Salverda, M. L. M. *et al.* Initial mutations direct alternative pathways of protein evolution. *PLoS Genet.* **7**, e1001321 (2011).

5. Hansen, T. F. & Wagner, G. P. Modelling genetic architecture: a multilinear theory of gene interaction. *Theor. Popul. Biol.* **59**, 61–86 (2001).

6. Breen, M. S., Kemena, C., Vlasov, P. K., Notredame, C. & Kondrashov, F. A. Epistasis as the primary factor in molecular evolution. *Nature* **490**, 535–538 (2012).

7. Li, W. H. *Molecular Evolution* (Sinauer, 1997).

8. da Silva, J. Site-specific amino acid frequency, fitness and the mutational landscape model of adaptation in HIV-1. *Genetics* **174**, 1689–1694 (2006).

9. Fowler, D. M. *et al.* High-resolution mapping of protein sequence-function relationships. *Nature Methods* **7**, 741–746 (2010).

10. Yang, Z. & Nielsen, R. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *J. Mol. Evol.* **46**, 409–418 (1998).

11. Choi, S. C., Redelings, B. D. & Thorne, J. L. Basing population genetic inferences and models of molecular evolution upon desired stationary distributions of DNA or protein sequences. *Phil. Trans. R. Soc. B* **363**, 3931–3939 (2008).

12. Rodrigue, N., Philippe, H. & Lartillot, N. Mutation-selection models of coding sequence evolution with site-heterogeneous amino acid fitness profiles. *Proc. Natl Acad. Sci. USA* **107**, 4629–4634 (2010).

13. Tamuri, A. U., dos Reis, M. & Goldstein, R. A. Estimating the distribution of selection coefficients from phylogenetic data using sitewise mutation-selection models. *Genetics* **190**, 1101–1115 (2012).

14. Yang, Z. PAML 4: a program package for phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).

15. Halpern, A. L. & Bruno, W. J. Evolutionary distances for protein-coding sequences: modeling site-specific residue frequencies. *Mol. Biol. Evol.* **15**, 910–917 (1998).

Competing Financial Interests Declared none.  
doi:10.1038/nature12219

Breen *et al.* reply

REPLYING TO D. M. McCandlish, E. Rajon, P. Shah, Y. Ding & J. B. Plotkin *Nature* **497**, <http://dx.doi.org/10.1038/nature12219> (2013)

Understanding fitness landscapes, a conceptual depiction of the genotype-to-phenotype relationship, is crucial to many areas of biology. Two aspects of fitness landscapes are the focus of contemporary studies of molecular evolution. First, the local shape of the fitness landscape defined by the contribution of individual alleles to fitness that is independent of all genetic interactions. Second, the global, multidimensional fitness landscape<sup>1</sup> shape determined by how interactions between

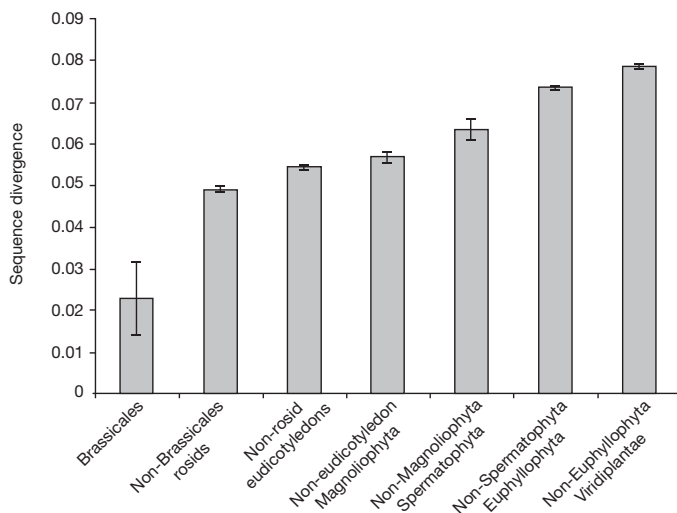
alleles at different loci change each other’s fitness impact, or epistasis. In explaining the high amino-acid usage ( $u$ ), we focused on the global shape of the fitness landscape<sup>2</sup>, ignoring the perturbations at individual sites<sup>3</sup>.  $u$  is overly sensitive to the presence of rare amino acids<sup>2</sup>, including fixed slightly deleterious amino acid—a factor we did not consider. McCandlish *et al.*<sup>3</sup> apply a non-epistatic model with fitness landscape



ruggedness on the local scale of individual sites. As a null model for testing the hypothesis of an epistasis-free fitness landscape, it directly confirms our conclusion that epistasis affects most amino-acid substitutions for 13 out of 16 genes considered<sup>3</sup>. The other three genes are extremely conservative with low density and frequency of emerged amino-acid states in the multiple sequence alignment. In the non-epistatic model<sup>3</sup> such amino-acid states seem to be substantially deleterious, leading to the low predicted dN/dS values in these three genes, with the largest effect in *rbcl*.

In the absence of epistasis, strong selection against non-optimal states markedly decreases the equilibrium sequence divergence<sup>3,4</sup> and the expected time to reach the equilibrium divergence<sup>4</sup>. The model<sup>3</sup> for *rbcl* simulates an equilibrium sequence divergence of ~0.06, which must be independent of phylogenetic distance beyond closely related clades<sup>4</sup>. Both of these predictions are easily falsified. Orthologous *rbcl* sequence divergence shows no sign of reaching a true equilibrium even between phylogenetically distant clades (Fig. 1), whereas a BLAST search reveals that sequence divergence between *Arabidopsis thaliana* and cyanobacterial orthologues reaches values greater than 0.16.

Generally, the non-epistatic model has a trade-off between the strength of selection against suboptimal alleles and the expected sequence divergence, which rapidly reaches its equilibrium value<sup>4,5</sup>.



**Figure 1 | Sequence divergence as a function of phylogenetic distance.** Average sequence divergence for pairwise comparisons of *A. thaliana* RbcL protein sequence and other orthologous sequences used in ref. 2.

For extremely conservative genes, such as the selected three non-mitochondrial genes considered<sup>2</sup>, the non-epistatic model can give the appearance of avoiding this trade-off, which breaks down when the long-term evolutionary predictions of the model are considered in detail (Fig. 1).

Two aspects of protein evolution are revealed by sequence similarity searches. First, protein sequence divergence occurs slowly, slower than neutral divergence. Second, sequence divergence is proportional to phylogenetic distance and is usually substantial for sequences from distantly related species. Non-epistatic models<sup>3,4</sup> that consider only local fitness landscape ruggedness are inconsistent with both of these basic and universal features of protein evolution. By contrast, our claim that epistasis—the global, multidimensional shape of the fitness landscape—is the primary factor of protein evolution explains the high amino-acid usage<sup>2</sup> and how slow long-term sequence divergence leads to highly dissimilar sequences<sup>5</sup>. Models that take into account both local and global aspects of fitness landscapes could lead to better quantification of factors shaping molecular evolution, although their development may be hampered by inherent complexity of multidimensional fitness landscapes<sup>1</sup> and subtle local confounding factors<sup>6–8</sup>.

**Michael S. Breen<sup>1</sup>, Carsten Kemena<sup>2,3</sup>, Peter K. Vlasov<sup>2,3</sup>, Cedric Notredame<sup>2,3</sup> & Fyodor A. Kondrashov<sup>2,3,4</sup>**

<sup>1</sup>Department of Medicine, University of California San Diego, La Jolla, California 92093, USA.

<sup>2</sup>Bioinformatics and Genomics Programme, Centre for Genomic Regulation (CRG) 88 Dr. Aiguader, 08003 Barcelona, Spain.

<sup>3</sup>Universitat Pompeu Fabra (UPF), 08003 Barcelona, Spain.

<sup>4</sup>Institució Catalana de Recerca i Estudis Avançats (ICREA), 23 Passeig Lluís Companys, 08010 Barcelona, Spain.

e-mail: fyodor.kondrashov@crge.es

- Kondrashov, F. A. & Kondrashov, A. S. Multidimensional epistasis and the disadvantage of sex. *Proc. Natl Acad. Sci. USA* **98**, 12089–12092 (2001).
- Breen, M. S., Kemena, C., Vlasov, P. K., Notredame, C. & Kondrashov, F. A. Epistasis as the primary factor in molecular evolution. *Nature* **490**, 535–538 (2012).
- McCandlish, D. M., Rajon, E., Shah, P., Ding, Y. & Plotkin, J. B. The role of epistasis in protein evolution. *Nature* **497**, E1–E2 (2013).
- Kondrashov, A. S., Povolotskaya, I. S., Ivankov, D. N. & Kondrashov, F. A. Rate of sequence divergence under constant selection. *Biol. Direct* **5**, 5 (2010).
- Povolotskaya, I. S. & Kondrashov, F. A. Sequence space and the ongoing expansion of the protein universe. *Nature* **465**, 922–926 (2010).
- McVean, G. & Charlesworth, B. A population genetic model for the evolution of synonymous codon usage: patterns and predictions. *Genet. Res.* **74**, 145–158 (1999).
- Kondrashov, F. A., Ogurtsov, A. Y. & Kondrashov, A. S. Selection in favor of nucleotides G and C diversifies evolution rates and levels of polymorphism at mammalian synonymous sites. *J. Theor. Biol.* **240**, 616–626 (2006).
- Kimura, M. The role of compensatory neutral mutations in molecular evolution. *J. Genet.* **64**, 7–19 (1985).

doi:10.1038/nature12220